

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:01:39 ; Search time 2272.12 Seconds

(without alignments)  
 363.693 Million cell updates/secTitle: US-10-085-612-2  
 Perfect score: 34  
 Sequence: 1 cggatctttgtcactggctgatcgatgcggccgg 34Scoring table: IDENTITY\_NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562/84

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : EST:\*

1: em\_estba:\*

2: em\_estbum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estsor:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htci:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estbn:\*

16: em\_estm:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fam:\*

22: em\_gss\_mus:\*

23: em\_gss\_pro:\*

24: em\_gss\_lod:\*

25: em\_gss\_phg:\*

26: em\_gss\_vrl:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
c 1	24.4	71.8	802	B2571326
c 2	23.4	68.8	910	CNS029D3
c 3	23	67.6	438	BMO3793
c 4	23	67.6	660	AA739600

FEATURES	Qualifiers
AW501182	UI-HF-BP0
BX82025	BX282025
AI56540	tnl0e09.x
B1868034	ft67ad7.y
BP131634	7c16f11.x
BG24809	602454595
BG335538	60240831
BG15582	60325515
BG911254	60280519
BI910096	603067939
BI154534	63181842
BG717956	60269932
BF944839	60230094
BG17765	60268226
BU54473	AGENCCOURT
CD398712	AGENCCOURT
AL26288	Tetradon
BU1933557	AGENCCOURT

## ALIGNMENTS

CC011462	PUDGZ331TD
AA03400	GN10776.5
AA538677	ID18261.5
BH613513	SALK_0343
BH14265	SALK_0104
T60367	ybb0095.rl
AJ478560	Au78560
AJ478560	Au78560
AJ478560	Au78560
BET30257	601533625
BJ219877	BJ219877
Al28.196	Tetradon
BG343301	HVSMB0000
B089835	AGENCCOURT
BW72531	AGENCCOURT
BM548921	AGENCCOURT
AI341794	Tetradon
BW005553	101111C0
BW621560	170006874
BW578845	170006871
AQ575468	nxv0087K
BH53181	SALK_0145
AW501182	UI-HF-BP0
BX82025	BX282025
AI56540	tnl0e09.x
B1868034	ft67ad7.y
BP131634	7c16f11.x
BG24809	602454595
BG335538	60240831
BG15582	60325515
BG911254	60280519
BI910096	603067939
BI154534	63181842
BG717956	60269932
BF944839	60230094
BG17765	60268226
BU54473	AGENCCOURT
CD398712	AGENCCOURT
AL26288	Tetradon
BU1933557	AGENCCOURT

## REFERENCE

AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.

Whole-Genome Sequence variation among multiple isolates of

## TITLE

J. Bacteriol.

(2002) In press

## JOURNAL

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062165054

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun

Location/Qualifiers

1. .802  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="msh2-1835"  
 /clone\_lid="msh"  
 /note="Environmental isolate. Whole genomic shotgun library."  
 BASE COUNT 139 a 218 c 164 g 228 t 53 others  
 ORIGIN

Query Match 71.8%; Score 24.4; DB 29; Length 802;  
 Best Local Similarity 82.4%; Pred. No. 1.6e+02;  
 Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Qy 1 CCAATTGCTTGTAGTCAGCTGGAGCTGCATCGAGGCCG 34  
 Db 348 CGATGGCTTGTCTAGGGCGCATCTGAGGCCG 315

RESULT 2  
 CNS02SD3/c  
 LOCUS CNS02SD3 910 bp DNA linear GSS 01-SEP-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
 161022 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 ACCESSION AL211728  
 GI:7870547  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetraodon nigroviridis  
 ORGANISM Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Neoreleosteii;  
 Tetradontoidae; Tetraodontidae; Tetraodontiformes;  
 Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fizames,C., Wincker,P., Brotlier,P., Quetier,F.,  
 Saurin,W. and Weissbach,J.  
 Estimate of human gene number provided by genome-wide analysis  
 using Tetraodon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)

REFERENCE 1  
 AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,  
 Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,  
 Saurin,W., Barnot,A. and Weissbach,J.  
 TITLE Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 JOURNAL Genom. Res. 10 (7), 939-949 (2000)  
 PMID 10835645  
 PUBMED 20296633  
 REFERENCE 2  
 AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,  
 Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,  
 Saurin,W., Barnot,A. and Weissbach,J.  
 TITLE Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 JOURNAL Genom. Res. 10 (7), 939-949 (2000)  
 PMID 10899143  
 PUBMED 2035837  
 REFERENCE 3  
 AUTHORS Genoscope,  
 Direct Submission  
 TITLE Submitted (12-APR-2000) Genoscope - Centre National de Séquençage :  
 BR 191 91006 EVRY cedex - FRANCE (E-mail : segrégogenoscope.ons.fr  
 - Web : www.genoscope.ons.fr)  
 COMMENT This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon/  
 Location/Qualifiers

FEATURES source  
 1. .910  
 /organism="tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3983"  
 /clone="1161722"  
 /clone\_lid="g"  
 /note="Genoscope sequence ID : COAG161DE11SP1-end :  
 PUC-Ori"

BASE COUNT 248 a 192 c 221 g 243 t 6 others  
 ORIGIN

Query Match 68.8%; Score 23.4; DB 29;  
 Best Local Similarity 81.8%; Pred. No. 3.7e+02;  
 Matches 27; Conservative 0; Mismatches 6;  
 Indels 0; Gaps 0;  
 Qy 2 GATCTTGACTGCTGAGCTGACCCG 34  
 Db 603 GATCTTGCTGAGGTGGCGTGCACCCG 571

RESULT 3  
 BN037193/c  
 LOCUS BN037193 438 bp mRNA linear EST 05-NOV-2001  
 DEFINITION fub4c01.y1 Gong zebrafish ovary Danio rerio cDNA clone  
 IMAGE:1537536 5, similar to SW:PMW2\_HUMAN 015305 PHOSPHOMANNOTASE  
 2 i, mRNA sequence.  
 ACCESSION BN037193  
 VERSION BN037193.1 GI:16750764  
 KEYWORDS EST,  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 438)  
 AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Boddy  
 ,S., Hillier,L., Kucaba,T., Martin,J.J., Beck,C., Wyllie,T., Underwood  
 ,K., Steptoe,M., Theising,B., Aihle,M., Bowers,Y., Person,B.,  
 Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
 Kohn,S., Shin,T.T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.,  
 and Wilson,R.  
 TITLE WashU Zebrafish EST Project 1998  
 JOURNAL Unpublished  
 COMMENT Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zebrafish@wustl.edu  
 The library was constructed by Dr. Z. Gong. DNA Sequencing by:  
 Washington University Genome Sequencing Center St. Louis. Please  
 contact Zhiyuan Gong for further information on this library  
 (National University of Singapore, Department of Biological  
 Sciences, Lower Kent Ridge Road, Singapore 119260).  
 High quality sequence stop: 323.  
 Location/Qualifiers

1. .438  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /clone\_xref="Taxon:7555"  
 /dev\_stage="embryo"  
 /lab\_host="DHL0B (phage-resistant)"  
 /clone\_lib="Gong zebrafish ovary"  
 /note="Organ: ovary (pooled); Vector: pBluescript SK-"  
 /Site\_1: Xhol; Site\_2: EcoRI; Poly A+ RNA was isolated from  
 the ovaries of 2 female adult zebrafish (4-5 month old).  
 cDNAs were made using oligo-dT primers and inserted into  
 lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo  
 mass-excised to pBluescript SK- following the Washington  
 University protocol.  
 ([http://genome.wustl.edu/est/lambda\\_protocol.shtml](http://genome.wustl.edu/est/lambda_protocol.shtml)).  
 Please contact Zhiyuan Gong for further information on  
 this library (National University of Singapore,  
 Department of Biological Sciences, Lower Kent Ridge Road,  
 Singapore 119260)."

BASE COUNT 138 a 78 c 100 g 122 t  
 ORIGIN  
 Query Match 67.6%; Score 23; DB 12; Length 438

Best Local Similarity 83.9%; Pred. No. 4.3e+02; Mismatches 26; Conservative 0; Indels 5; Gaps 0; Reference 1 (bases 1 to 704).  
 Clade: Panicoideae; Andropogoneae; Zea.  
 Whitehall C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick A., Frazer, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.  
 Maize Genomics Consortium  
 Unpublished  
 Contact: Cathy Whitelaw

RESULT 4  
 LOCUS AA739600 660 bp mRNA taeda cDNA clone 8604M 3', mRNA sequence.  
 DEFINITION AA739600.1 GI:2778151  
 VERSION AA739600.1  
 KEYWORD EST.  
 ORGANISM Pinus taeda (loblolly pine)  
 SOURCE Eukaryota; Viriplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus 1 (bases 1 to 660)  
 REFERENCE Kinlaw,C.S.  
 AUTHORS Loblolly pine cDNAs  
 JOURNAL Unpublished  
 COMMENT Contact: Claire S. Kinlaw  
 USDA IRG Dendrome Project  
 Institute of Forest Genetics  
 Dendrome Project, Institute of Forest Genetics, P.O. Box 245,  
 Berkeley, CA 94701  
 Tel: 5105596429  
 Fax: 5105596440  
 Email: csk@27w007.pswfs.gov  
 The sequence entry for this EST has been reverse complimented and is being submitted to MI3 Universal.  
 Seq primer: MI3.Universal  
 Location/Qualifiers 1. .660  
 /organism="Pinus taeda"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:33532"  
 /clone="8604M"  
 /tissue\_type="xylem"  
 /clone\_id="PTFIG21"  
 /note="Vector: Lambda ZAP; Site\_1: EcoRI; Site\_2: XbaI;  
 The tissue source for this library is xylem. The xylem tissue was harvested in spring and summer from branches of seed orchard trees which are clones of the same genotype. Branches were 4-6 inches in diameter. The cDNAs were directionally cloned into Lambda Zap and were rescued as a Bluescript derivative in the EcoRI and XbaI sites."  
 BASE COUNT 201 a 135 c 108 g 215 t 1 others  
 ORIGIN Query Match 67.6%; Score 23; DB 9; Length 660;  
 Best Local Similarity 83.9%; Pred. No. 4.8e+02; Mismatches 26; Conservative 0; Indels 5; Gaps 0; Reference 1 (bases 1 to 704).  
 Clade: Zea mays genomic clone ZMBBta185E14, genomic survey sequence.  
 CC011462.1 GI:23392738  
 GSS  
 KEYWORDS Zea mays  
 SOURCE Eukaryota; Viriplantae; Streptophyta; Embryophyta; Tracheophyta;  
 ORGANISM Maize Genomics Consortium  
 Unpublished  
 Contact: Cathy Whitelaw

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TP  
 Class: sheared ends.  
 FEATURES Source  
 /Location/Qualifiers 1. .704  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMTRA15E14"  
 /clone\_id="ZM\_0\_6\_1\_0\_KB"  
 /note="Vector: PCR4-TOPO; Site\_1: BCGRI; 0.6-1.0 kb high Cot selected genomic DNA library"  
 BASE COUNT 151 a 181 c 230 g 142 t  
 ORIGIN Query Match 66.5%; Score 22.6; DB 29; Length 704;  
 Best Local Similarity 86.2%; Pred. No. 6.7e+02; Mismatches 25; Conservative 0; Indels 0; Gaps 0;  
 Matches 25  
 QY 6 CTTGGTACTGCCTGAGGTGCAACCCCG 34  
 DB 294 CTCCTCTCGAGGCTGAACTGCAACCCCG 266

RESULT 6  
 AA803400 603 bp mRNA linear EST 23-APR-2001  
 LOCUS AA803400  
 DEFINITION GM10776.5prime GM Drosophila melanogaster ovary BlueScript to  
 Drosophila melanogaster cDNA clone GM10776 5prime similar to  
 X78555: CalPA FBgn012051 PID:19562289 SWISS-PROT:Q11002, mRNA  
 sequence.  
 ACCESSION AA803400  
 VERSION AA803400.1 GI:28772519  
 KEYWORD Drosophila melanogaster (fruit fly)  
 SOURCE Drosophila melanogaster  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydioidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 603).  
 AUTHORS Harvey,D., Brookstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.  
 TITLE BDGP/HMM Drosophila EST Project  
 JOURNAL Unpublished  
 COMMENT Contact: Stapleton, M.  
 BDGP Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Email: http://www.fruitfly.org/EST/est@fruitfly.berkeley.edu  
 Fax: 510 486 6798  
 Plate: 107 row: G column: 4  
 High quality sequence stop: 486.  
 Location/Qualifiers 1. 603  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clones="Q110776"  
 /sex="female"  
 /dev\_stage="newly eclosed females: germarium-stage 6"

RESULT 5  
 CC011462/C 704 bp DNA linear GSS 31-MAR-2003  
 LOCUS CC011462 DEFINITION PUDG211TD ZM\_0\_6\_1.0\_KB zeamays genomic clone ZMBBta185E14,  
 ACCESSION CC011462  
 VERSION CC011462.1 GI:23392738  
 KEYWORDS GSS  
 SOURCE Zea mays  
 ORGANISM Maize Genomics Consortium  
 Unpublished  
 Contact: Cathy Whitelaw

BASE COUNT	139	a	154	c	138	t	ORIGIN	Qy	1 CGATTCTTGACTACGGCTGCAGCTGC 27                                     /clone lib="SOLR"	Db	545 CGATTGTTGGTACTGGCTGCAGCTGC 571                                 /clone lib="SOLR"
DEFINITION								Qy	BH613513	BH613513	
LOCUS								Db	SALK_034371	SALK_034371	
VERSION								Qy	Arabidopsis thaliana	Arabidopsis thaliana	
KEYWORDS								Db	(thale cress)	(thale cress)	
SOURCE								Qy	Eukaryota; Viridiplanteae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; rosids	Eukaryota; Viridiplanteae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; rosids	
ORGANISM								Db	eurosid II; Brassicales; Brassicaceae; Arabidopsis	eurosid II; Brassicales; Brassicaceae; Arabidopsis	
REFERENCE								Qy	1 (bases 1 to 180) / Zimmermann, J. and Ecker, J.R. / C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Precinis, L., Shinn, P.	1 (bases 1 to 180) / Zimmermann, J. and Ecker, J.R. / C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Precinis, L., Shinn, P.	
AUTHORS								Db	Unpublished	Unpublished	
TITLE								Qy	Arabidopsis thaliana	Arabidopsis thaliana	
JOURNAL								Db	COMMENT	COMMENT	
COMMENT								Qy	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL)	Contact: Joseph R. Ecker Salk Institute for Biological Studies	
								Db	The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA	The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA	
								Qy	Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu	Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu	
								Db	CLASS: TDNA tagged	CLASS: TDNA tagged	
								Qy	This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of Atig20480.	This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of Atig20480.	
RESULT	7							Qy	1..180	1..180	
LOCUS	AA538677							Qy	/organism="Arabidopsis thaliana" /mol_type="genomic DNA" /strain="Columbia 0" /db_xref="taxon:3702"	/organism="Arabidopsis thaliana" /mol_type="genomic DNA" /strain="Columbia 0" /db_xref="taxon:3702"	
DEFINITION	LD18261.5prime LD Drosophila melanogaster embryo BlueScript							Qy	/clone.lib="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://sign.salk.edu/tDNA_protocols.html"	/clone.lib="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://sign.salk.edu/tDNA_protocols.html"	
ORGANISM	Drosophila melanogaster							Qy	BASE COUNT	BASE COUNT	
VERIFICATION	EST.							Qy	ORIGIN	ORIGIN	
KEYWORDS	Drosophila melanogaster (fruit fly)							Qy	Query Match	Query Match	
SOURCE	Drosophila melanogaster							Qy	Best Local Similarity	Best Local Similarity	
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;							Qy	Score	Score	
VERIFICATION	EST.							Qy	22; DB 28;	22; DB 28;	
KEYWORDS	Drosophila melanogaster							Qy	Length	Length	
SOURCE	Drosophila melanogaster							Qy	180;	180;	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	Best Matches	Best Matches	
VERIFICATION	EST.							Qy	83.3%; Pred.	83.3%; Pred.	
KEYWORDS	Drosophila melanogaster							Qy	No. 7.9e+02;	No. 7.9e+02;	
SOURCE	Drosophila melanogaster							Qy	Mismatches	Mismatches	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	5; Indels	5; Indels	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
ORGANISM	Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;							Qy	0; Gaps	0; Gaps	
VERIFICATION	EST.							Qy	0; Gaps	0; Gaps	
KEYWORDS	Drosophila melanogaster							Qy	0; Gaps	0; Gaps	
SOURCE	Drosophila melanogaster							Qy</			



Plate: LICM511 row: b column: 08  
 High quality sequence stop: 667.  
 Location/Qualifiers  
 1..667  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3832999"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone.lib="NIH\_MGC\_20"  
 /note="Orgn: skin; Vector: pOTB7; Site\_1: XbaI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XbaI sites using the following 5'  
 adaptor: GCGCCAGG(G). Site selected >50bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

**RESULT 12**

**LOCUS** AJ432574 509 bp tRNA linear EST 15-NAR-2002  
**DEFINITION** AJ432574 S00011 Hordeum vulgare cDNA clone S0001100156E11F1, mRNA  
**SEQUENCE**.

**ACCESSION** AJ432574  
**VERSION** GI:19521026  
**KEYWORDS**  
**ORGANISM** Hordeum vulgare  
 Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Pooidae  
 ; Triticeae; Hordeum.  
**REFERENCE**  
**AUTHORS** Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.  
**TITLE** Barley EST's  
**JOURNAL** Unpublished  
**COMMENT** Contact: Schulman AH  
 Institute of Biotechnology  
 University of Helsinki  
 P.O.Box 56 (Viikinkaari 6A); University of Helsinki FIN-00014,  
 Finland.

**FEATURES**  
**source**

1..509  
 /organism="Hordeum vulgare"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4513"  
 /clone="S0001100156E11F1"  
 /dev\_stage="Developing seed"  
 /clone.lib="S00011"  
 /note="12,15,18 days after pollination"

**BASE COUNT** 104 a 142 g 87 t  
**ORIGIN**

Query Match 64.1%; Score 21.8; Length 509;  
 Best Local Similarity 78.8%; Pred. No. 1.2e+03  
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

**ACCESSION** QY 2 GATTCTTGACTGCTGAGTCAGCTGCCGG 34  
**VERSION** Db 77 GATTGCCCTCAGGTCTCGTCAGGCCCG 109

**FEATURES**  
**source**

RESULT 14  
 BTJ19877/C  
**LOCUS** BTJ19877 687 bp mRNA linear EST 04-APR-2002  
**DEFINITION** BTJ19877 Y. Oghara unpublished cDNA library, Wh Triticum aestivum  
 cDNA clone wh17g09 3', mRNA sequence.

**ACCESSION** BTJ19877  
**VERSION** BTJ19877.1 GI:19958527  
**KEYWORDS**  
**ORGANISM** Triticum aestivum (bread wheat)  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyt; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
**REFERENCE** 1 (bases 1 to 687)  
**AUTHORS** Oghara,Y. and Murai,K.  
**TITLE** Unpublished  
**JOURNAL**  
**COMMENT** Contact: Tadasu Shin-i  
 Center for Generic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
**FEATURES**  
**source**

1..687  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="wh17g09"  
 /tissue\_type="spike at meiosis"  
 /dev\_stage "Feeke's" scale 9"  
 /clone.lib="Y. Oghara unpublished cDNA library, Wh"

**BASE COUNT** 174 a 203 g 147 t  
**ORIGIN**

Query Match 64.1%; Score 21.8; DB 12; Length 687;  
 Best Local Similarity 78.8%; Pred. No. 1.3e+03;  
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

**ACCESSION** QY 2 GATTCTTGACTGCTGAGTCAGCTGCCGG 34



GenCore version 5.1.6  
Copyright (C) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:01:39 ; Search time 1202.88 Seconds  
(without alignments)  
363.693 Million cell updates/sec

Title: US-10-085-612-1  
Perfect score: 18  
Sequence: 1 gacaaggccaggcagacag 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12125238056 residues

Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries  
Database : EST:\*

```
1: em_estba:*
2: em_estbin:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estin:*
16: em_eston:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_fut:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_pro:*
24: em_gss_rod:*
25: em_gss_phg:*
26: em_gss_yrl:*
27: em_gss_yrl:*
28: gb_gss1:*
29: gb_gss2:*
```

RESULT 1  
AZ722759

LOCUS

DEFINITION

ACCESSION AZ722759

VERSION AZ722759.1

KEYWORDS GSS,

SOURCE Mus musculus (house mouse)

ORGANISM

Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Other\_GSS: RPCI-24-144M13-TV

COMMENT

REFERENCE

AUTHORS 1. (bases 1 to 546)  
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregiorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse\_BAC End Sequences from Library RPCI-24

TITLE

JOURNAL

ALIGNMENTS

RESULT 1  
BY115518  
LOCUS BY115518  
DEFINITION linear DNA  
genomic survey sequence.

ACCESSION BY20897

VERSION BY495987

KEYWORDS EST05216

SOURCE BG777302

ORGANISM AMGNNNUC:C

CB771456

BY627146

BY277146

BP117896

BP117896

BB761925

BB761925

BB20897

UL-B-EPO-

BQ20897

EST05216

BG711834

&lt;p

library availability, please contact Pieter de Jong (pdejong@mail.chb.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tgb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tgb/bac_ends/mouse/bac_end_intro.html)

Plate: 144 row: M column: 13

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 546 /organism="Mus musculus" /mol\_type="genomic DNA" /strain="C57BL/6J" /db\_xref="Taxon:10090" /clone="RPCI-24-14A13" /sex="Male" /cell\_type="Spleen/Brain" /clone\_lib="RPCI-24"

/note="Vector: PTARACI; Site\_1: BamH1; Site\_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the PTARACI cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 158 a 117 c 119 g 152 t

ORIGIN Query Match 100.0% Score 18; DB 28; Length 546; Best Local Similarity 100.0% Pred. No. 1.1e+03; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1. GACAGCCGAGCACAG 18  
Db 92 GACAGGGCAGGACAG 109

RESULT 2 BG020388/c LOCUS dc73g11.x1 NICHD\_XGC\_Emb1 Xenopus laevis mRNA linear EST 24-JAN-2001 DEFINITION 3' mRNA sequence.

VERSION BG020388 EST.

KEYWORDS Xenopus laevis (African clawed frog)

SOURCE Xenopus laevis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus.

REFERENCE NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), tumor Gene Index

TITLE Unpublished

JOURNAL Contact: Robert Strausberg, Ph.D.  
Email: cgaps@mail.nih.gov  
Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LNL at: [info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: 40UP from Gibco  
High quality sequence stop: 412.

Location/Qualifiers

1. 487 /organism="Xenopus laevis" /mol\_type="mRNA" /db\_xref="taxon:8355" /clone="IMAGE:3402836" /tissue\_type="embryo (stage 10)" /lab\_host="DH10B (pGEM-resistant)" /clone\_lib="NICHD\_XGC\_Emb1"  
/note="Vector: pCMV-SF0R6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert

FEATURES source

size 1.55 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library." BASE COUNT 159 a 102 c 87 g 139 t

ORIGIN Query Match 94.4%; Score 17; DB 10; Length 487; Best Local Similarity 100.0%; Pred. No. 2.8e+03; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ACAAGGGAGGACAGAG 18  
Db 251 ACAAGGGAGGACAGAG 235

RESULT 3 BH305292/c LOCUS BH305292 TV CHORI-230 Segment 1 Rattus norvegicus genomic clone DEFINITION CH230-100K7, CH230-100K7, genomic survey sequence.

ACCESSION BH305292 VERSION BH305292.1 GI:17217700 GSS:

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus (Norway rat) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.

REFERENCE 1. (bases 1 to 647) Zho,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn A., Gabregeorgis,P., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong, P. and Fraser,C.M.

AUTHORS Rat BAC End Sequences from Library CHORI-230 ECO RI segment Unpublished

TITLE JOURNAL Other\_GSS: CH230-100K7.TU

COMMENT Contact: Shaying Zhao Email: [szhao@tigr.org](mailto:szhao@tigr.org) Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.chb.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering\_information.htm). BAC end page: http://www.tigr.org/tgb/bac\_ends/rat/bac\_end\_intro.html Plate: 100 row: K column: 7 Seq Primer: T7 Class: BAC ends.

FEATURES source Location/Qualifiers

1. .647 /organism="Rattus norvegicus" /mol\_type="genomic DNA" /strain="BN/SsNhsd/MCW" /db\_xref="CH230-100K7" /clone="CH230-100K7" /sex="Female"

/cell\_type="Brain"

/clone\_id="CHORI-230 Segment 1"

/note="Vector: PTARBA2.1; Site\_1: EcoRI; Site\_2: EcoRI; CHORI-230 Rat (BN/SsNhsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 157 a 143 c 193 g 154 t

ORIGIN Query Match 94.4%; Score 17; DB 28; Length 647; Best Local Similarity 100.0%; Pred. No. 3e+03; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ACAAGGGAGGACAGAG 18  
Db 17 ACAAGGGAGGACAGAG 1





Db	148	GACAAGGGCAGCACAGAG	165												
RESULT	7														
LOCUS	BG004130	CM2-GN0166-201100-525-g11	207 bp	mRNA	linear	EST 24-JAN-2001									
DEFINITION				Homo sapiens	cDNA,	mRNA sequence.									
VERSION	BG004130.1	GI:12444997													
KEYWORDS															
ORGANISM	Homo sapiens	(human)													
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.															
REFERENCE	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,M.A., Bordim,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsuoka,A., Boia,G.S., Simpson,D.H., Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.														
AUTHORS															
COMMENT	1 ("bases 1 to 207")														
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)														
FEATURES															
SOURCE															
COMMENT	1. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
COMMENT	2. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	3. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	4. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	5. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	6. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	7. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	8. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	9. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	10. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	11. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	12. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	13. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	14. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	15. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	16. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	17. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	18. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	19. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	20. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	21. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	22. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	23. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	24. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	25. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	26. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	27. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	28. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	29. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	30. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	31. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	32. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	33. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	34. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	35. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	36. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	37. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES															
SOURCE															
COMMENT	38. 245 /organism="Salmo salar" /mol-type="mRNA" /strain="McConnell" /db_xref="taxon:8030"														
FEATURES				</											



3']. cDNA was cloned into the XbaI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

BASE COUNT 71 a ORIGIN 79 c 50 g 73 t

Query Match 91.1%; Score 16.4; DB 10; Length 273;  
Best Local Similarity 94.4%; Pred. No. 4.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGGAGAG 18  
||| | | | | | | | | | | | | |  
Db 145 GAGAGGGCAGGAGAG 128

RESULT 11.

AZ902923 LOCUS AZ902923 277 bp DNA linear GSS 05-MAR-2001  
DEFINITION RPCI-24-155K10.TJ RPCI-24 Mus musculus genomic clone RPCI-24-155K10  
ACCESSION AZ902923  
VERSION AZ902923.1 GI:13221968  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Eutheria; Rodentia; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathii; Murinae; Murinae; Mus.

REFERENCE 1 (bases 1 to 277)  
AUTHORS Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akinrot,B., Levin,M.,  
Tsegaye,G., Geer,K., Krolik,M., Shvartsbeyn,A., Gebregeorgis,E.,  
Russell,D., de Jong,P. and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24  
JOURNAL Unpublished  
COMMENT Other GSSs: RPCI-24-155K10.TV  
Contact: Shzao zhao  
Deparment of Bukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@email.cho.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/rdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/rdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 155 row: K column: 10  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers

FEATURES source  
1. 277 /organism="Mus musculus"  
/clone\_type="Spliced/Brain"  
/clone\_lib="RPCI-24"  
/note="Vector: pBAC1; Site\_1: BamH1; Site\_2: BamH1;  
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 63 a ORIGIN 84 c 83 g 47 t

Query Match 91.1%; Score 16.4; DB 28; Length 277;  
Best Local Similarity 94.4%; Pred. NO. 4.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGGAGAG 18  
||| | | | | | | | | | | | | |

226 GACAGGGCAGGAGAG 209

Db RESULT 12  
LOCUS BB006387/C  
DEFINITION BB006387 RIKEN full-length enriched, 10 day neonate skin Mus  
musculus cDNA clone 473247F02 3', mRNA sequence.  
ACCESSION BB006387  
VERSION EST.  
KEYWORDS GI:8095786  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 288)  
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
Hirozane,T., Hori,F., Ishii,Y., Itoh,I., Itoh,M., Itoh,T.,  
Izawa,M., Kadota,K., Kawai,T., Kikuchi,N.,  
Kiyosawa,H., Koizumi,Y., Kondo,S., Roya,S., Kurihara,C., Kusakabe,M.,  
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K.,  
Shigenobu,Y., Shinagawa,A., Shiraki,T., Sugahara,Y.,  
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Toninaga,N., Toya  
T., Tsunoda,Y., Watanabe,S., Yamamura,T., Yamanaka,I.,  
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshikli,A., Yoshino  
M., Muramatsu,M. and Hayashizaki,Y.  
TITLE RIKEN Mouse ESTs (Konno,H., et al.)  
JOURNAL Unpublished  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN) 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
The Institute for Physical and Chemical Research (RIKEN) Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp/  
URL:<http://genome.gsc.riken.go.jp/>  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., NakaoKA,S., Sasaki  
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermoactivation of thermostable enzymes by trehalose and its application to the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome rtc.riken.go.jp>) for further details.

FEATURES source  
1. • 288 /organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="#Taxon:10090"  
/sex="mixed"  
/tissue\_type="skin"  
/dev\_stage="10 days neonate"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 10 day neonate skin,"

/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was

RESULT	13	AZ464111/C	AZ464111	DNA	312 bp	DNA	linear	GSS 04-OCT-2000
LOCUS		IM0273A19F	Mouse 10kb plasmid UGCG1M library	Mus musculus	genomic clone	UNGCG1M0273A19 F,	genomic survey sequence.	
DEFINITION		AZ464111						
ACCESSION		AZ464111.1						
VERSION		GSS						
KEYWORDS		AZ464111.1 GI:10622236						
SOURCE		Mus musculus	(house mouse)					
ORGANISM		Mus musculus	(house mouse)					
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
AUTHORS		Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islan,R., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.						
TITLE		Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts						
JOURNAL		Unpublished						
COMMENT		Contact: Robert B. Weiss						
		University of Utah Genome Center						
		University of Utah						
		Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA						
		tel : 801 585 5606						
		Fax : 801 585 7177						
		Email: dmann@genetics.utah.edu						
		Insert Length: 10000 Std Error: 0.00						
		Plate: 0273 Row: A column: 19						
		Seq primer: CGTGTAAACGAGGCCAGT						
		Class: plasmid ends						
		High quality sequence stop: 312.						
FEATURES		Location/Qualifiers						
		1. .312						
		/organism="Mus musculus"						
		/mol_type="genomic DNA"						
		/strain="C57BL/6J"						
		/db_xref="taxon:10090"						
		/clone="UGCG1M0273A19"						
		/sex="Male"						
		/lab_host="E. Coli strain X110-Gold, T1-resistant, F-"						
		/clone_id="Mouse 10kb plasmid UGCG1M library"						
		/note="Vector: pMD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource ( <a href="http://www.jax.org/resources/documents/dnares/">http://www.jax.org/resources/documents/dnares/</a> ). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch oriifce at constant velocity. The sheared DNA was blunt end-repaired with M4 DNA polymerase and M4						

polymerase kinase. Adaptor Oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.<sup>11</sup>

RESULT	15	BY615518	349 bp mRNA	EST 15-DEC-2002
DEFINITION	BY615518 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K330010005 3'			
ACCESSION	BY615518			
VERSION	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus. 1 (bases 1 to 349)			
AUTHORS	Nikaido, Y., Furuno, M., Kasukawa, T., Adachi, J., Suzuki, H., Kondo, S., Osoato, N., Saico, R., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Battalov, S., Beisel, R.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani L.E., Cousins, S., Daila, E., Dragani, T.A., Forrest, A., Frazer, K.S., Gaasterland, T., Garibaldi, M., Gissi, C., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kavaji, H., Kawasawa, Y., Kejzralski, R.M., King, B.J., Kongaya, A., Kuroochi, J.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nunata, K., Okido, T., Pavani, W.J., Peretea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ranachandra, S., Ravasi, T., Reed, J.C., Reid, J.J., Reid, J.J., Ring B.Z., Rinwaal, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teesdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanebe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yang, C., Yang, L., Yang, Z., Zarولan, M., Zhu, Y., Zimmer, A., Canini, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazuni, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashimoto, K., Imoto, K., Iishi, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibusawa, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Rogers, J., Birney, E., and Hayashizaki, Y.			
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs			
TOPIC	Nature, 420, 553-557 (2002)			

22254683  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
tel: 81-45-503-9222  
fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane-Kishikawa, T., Itoh, M., Konno, H., Miyazaki, A., Motani, K., Iishi, Y., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Murata, M., Nakamura, M., Sasada, D., Sato, K., Shibata, K., Tagami, N., Sakazume, N., Sasaki, D., Shiraki, T., Waki, K., Watanuki, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome*. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in RIKEN. Genomic Sciences Center and Genome Science Laboratory in RIKEN.

revision of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Michela Fagiolini and Takao K. Hensch ( Laboratory for Neuronal Circuit Development Brain Science Institut RIKEN 2-1 Hirosawa Wako-shi Saitama 351-0198 Japan ) whose assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES	source	BASE COUNT	ORIGIN
Location/Qualifiers			
1..349			
/organism="Mus musculus"			
/mol_type="mRNA"			
/strain="C57BL/6J"			
/db_xref="TAXON:10090"			
/clone="K3001005"			
/tissue_type="visual cortex"			
/clone.lib="TRINN full-length enriched, visual cortex"			
90 a	98 c	82 g	79 t

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:42:00 ; Search time 91.7308 Seconds  
 (without alignments) 450.193 Million cell updates/sec

Title: US-10-085-612-1  
 Perfect score: 18 1 gacaaggccggacaaqg 18

Scoring table: IDENTITY\_NUC Gapop 10.0 , Gapext 1.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Published\_Applications\_NA:  
 1: /cgns2\_6/podata/2/pubpna/US07\_PUBCOMB.seq:  
 2: /cgns2\_6/podata/2/pubpna/US06\_PUB.seq:  
 3: /cgns2\_6/podata/2/pubpna/US05\_PUB.seq:  
 4: /cgns2\_6/podata/2/pubpna/US06\_PUBCOMB.seq:  
 5: /cgns2\_6/podata/2/pubpna/BCTUS\_PUBCOMB.seq:  
 6: /cgns2\_6/podata/2/pubpna/US09\_PUBCOMB.seq:  
 7: /cgns2\_6/podata/2/pubpna/US09\_NEW\_PUB.seq:  
 8: /cgns2\_6/podata/2/pubpna/US10\_PUBCOMB.seq:  
 9: /cgns2\_6/podata/2/pubpna/US09A\_PUBCOMB.seq:  
 10: /cgns2\_6/podata/2/pubpna/US09B\_PUBCOMB.seq:  
 11: /cgns2\_6/podata/2/pubpna/US09C\_PUBCOMB.seq:  
 12: /cgns2\_6/podata/2/pubpna/US09\_NEW\_PUB.seq:  
 13: /cgns2\_6/podata/2/pubpna/US10A\_PUBCOMB.seq:  
 14: /cgns2\_6/podata/2/pubpna/US10B\_PUBCOMB.seq:  
 15: /cgns2\_6/podata/2/pubpna/US10\_NEW\_PUB.seq:  
 16: /cgns2\_6/podata/2/pubpna/US60\_NEW\_PUBCOMB.seq:  
 17: /cgns2\_6/podata/2/pubpna/US60\_PUBCOMB.seq:

RESULT 1  
 US-10-085-612-1  
 ; Sequence 1, Application US/10085612  
 ; Publication No. US0030096251A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Guida, Marco  
 ; APPLICANT: Hall, Jeff  
 ; APPLICANT: Petros, William  
 ; APPLICANT: Vredenburg, James  
 ; APPLICANT: Colvin, Oliver  
 ; APPLICANT: Marks, Jeffrey  
 ; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals  
 ; FILE REFERENCE: 4389-5-CI  
 ; CURRENT APPLICATION NUMBER: US/10/085,612  
 ; PRIORITY FILING DATE: 2002-02-16  
 ; PRIORITY APPLICATION NUMBER: 09/144,367  
 ; PRIORITY FILING DATE: 1998-08-31  
 ; PRIORITY APPLICATION NUMBER: 60/271,630  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SEQ ID NO: 1  
 ; LENGTH: 18  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-085-612-1  
 ; Query Match 100.0% ; Score 18; DB 14; Length 18;  
 ; Best Local Similarity 100.0%; Pred. No. 18;  
 ; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query ID	Match ID	Length	DB	Description
1	18	100.0	18	14	US-10-085-612-1	Sequence 1, Appli
c 2	16.4	91.1	29	14	US-10-085-612-1	Sequence 13, Appli
c 3	16.4	91.1	529	13	US-10-027-632-105890	Sequence 10680, Appli
c 4	16.4	91.1	714	13	US-10-027-632-25559	Sequence 25159, Appli
c 5	16.4	91.1	1530	14	US-10-198-446-10434	Sequence 10434, Appli
c 6	16.4	91.1	119596	14	US-10-270-3346-3	Sequence 3, Appli
c 7	16.4	91.1	172637	9	US-09-805-458A-3	Sequence 3, Appli
c 8	16	88.9	745	13	US-10-027-632-165368	Sequence 166368, Appli
c 9	16	88.9	1345	14	US-10-146-575-3	Sequence 3, Appli
c 10	16	88.9	1345	14	US-10-085-612-3	Sequence 3, Appli
c 11	16	88.9	3881	11	US-09-848-820-20	Sequence 44, Appli
c 12	15.4	85.6	19	14	US-10-146-575-44	Sequence 3955, Appli
c 13	15.4	85.6	299	10	US-09-783-590-3955	Sequence 3955, Appli
c 14	15.4	85.6	591	13	US-10-027-632-322713	Sequence 321713, Appli
c 15	15.4	85.6	592	13	US-10-027-632-321714	Sequence 321714, Appli
c 16	15.4	85.6	592	13	US-10-027-632-72506	Sequence 72506, Appli

RESULT 2  
 US-10-146-575-13/c

Page

RESULT 6 US-10-270-336-3  
 Sequence 3, Application US/10270336  
 Publication No. US20030074673A1  
 GENERAL INFORMATION:  
 ; APPLICANT: ZHU, Shiaoping et al.  
 ; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF  
 ; FILE REFERENCE: CL001146CON  
 ; CURRENT APPLICATION NUMBER: US/10/270,336  
 ; CURRENT FILING DATE: 2002-10-15  
 ; PRIOR APPLICATION NUMBER: US 60/270,873  
 ; PRIOR FILING DATE: 2001-02-26  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 3 LENGTH: 119596  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE: misc\_feature  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1) ..(115956)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-10-270-336-3

Query Match 91.1%; Score 16.4; DB 14; Length 119596;  
 Best Local Similarity 94.4%; Pred. No. 46;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18  
 Db 63483 GACAAGGGCAGGAGAG 63500

---

RESULT 7 US-09-805-458A-3  
 Sequence 3, Application US/09805458A  
 ; Patent No. US200402100A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAN, Chunhua et al.  
 ; TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS, AND USES THEREOF  
 ; FILE REFERENCE: CL000722  
 ; CURRENT APPLICATION NUMBER: US/09/805,458A  
 ; CURRENT FILING DATE: 2001-03-14  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 3 LENGTH: 172637  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE: misc\_feature  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1) ..(172637)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-09-805-458A-3

Query Match 91.1%; Score 16.4; DB 9; Length 172637;  
 Best Local Similarity 94.4%; Pred. No. 44;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

---

RESULT 8 US-10-027-632-166368/C  
 Sequence 166368, Application US/10027632

Query Match 91.1%; Score 16.4; DB 9; Length 172637;  
 Best Local Similarity 94.4%; Pred. No. 44;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18  
 Db 95503 GACAAGGGCAGGAGAG 95520

---

RESULT 9 US-10-027-632-166368  
 Sequence 3, Application US/10027632  
 ; SEQ ID NO. 166368  
 ; SOFTWARE: FastSEQ for Windows Version 4.0

Query Match 88.9%; Score 16; DB 14; Length 1345;  
 Best Local Similarity 88.9%; Pred. No. 1..le0+2;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18  
 Db 806 GACAAGGGCAGGAGAG 823

---

RESULT 10 US-10-085-612-3

RESULT 11  
 Sequence 3, Application US/10085612  
 Publication No. US2003009625A1  
 GENERAL INFORMATION:  
 ; APPLICANT: Guida, Marco  
 ; APPLICANT: Hall, Jeff  
 ; APPLICANT: Petros, William  
 ; APPLICANT: Vredenburg, James  
 ; APPLICANT: Colvin, Oliver  
 ; APPLICANT: Marks, Jeffrey  
 ; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals  
 ; FILE REFERENCE: 4389-5-C1  
 ; CURRENT APPLICATION NUMBER: US/10/085,612  
 ; CURRENT FILING DATE: 2002-02-26  
 ; PRIOR APPLICATION NUMBER: 09/144,367  
 ; PRIOR FILING DATE: 1998-08-31  
 ; PRIOR APPLICATION NUMBER: 60/271,630  
 ; PRIOR FILING DATE: 2001-02-26  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO: 3  
 ; LENGTH: 1345  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-085-612-3

Query Match  
 Best Local Similarity 88.9%; Score 16; DB 14; Length 1345;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGAGAGACAGAG 18  
 Db 1|||||1|||1|||1|||1|||1|||  
 806 GACAAGGAGAGACAGAG 823

RESULT 13  
 Sequence 13, Application US-793-590-3955  
 ; Sequence 3555, Application US/09783590  
 ; PATENT NO. US2002110850A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dillon, Patrick J.  
 ; APPLICANT: Haseltine, William A.  
 ; APPLICANT: Li, Heding  
 ; APPLICANT: Rosen, Craig A.  
 ; APPLICANT: Ruben, Steven M.  
 ; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
 ; FILE REFERENCE: PO-16.2C1  
 ; CURRENT APPLICATION NUMBER: US/09//783,590  
 ; PRIOR APPLICATION NUMBER: 08/215  
 ; PRIOR FILING DATE: 2000-02-15  
 ; PRIOR APPLICATION NUMBER: 08/320,856  
 ; PRIOR FILING DATE: 1995-04-12  
 ; PRIOR APPLICATION NUMBER: 08/346,731  
 ; PRIOR FILING DATE: 1994-11-21  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 3955  
 ; LENGTH: 299  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (2)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (8)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (15)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (23)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (45)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (123)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (131)

RESULT 12  
 Sequence 12, Application US-10-575-44  
 ; Sequence 44, Application US/10146575  
 ; Publication No. US2003005980A1  
 ; GENERAL INFORMATION:

```

OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (113)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (204)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (214)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (256)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (271)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (294)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-3955

Query Match          85.6%; Score 15.4; DB 10; Length 299;
Best Local Similarity 88.9%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
; SEQ ID NO: US-10-027-632-286257/C

Qy      1 GACAGGGCAGCACAG 18
        ||| | | | | | | | | |
Db      39 GAAAGGNGCAGGACAG 56

RESULT 14
US-10-027-632-286257/C
; Sequence 286257, Application US/10027632
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-13
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO: 321713
;   LENGTH: 592
;   TYPE: DNA
;   ORGANISM: Human
;   US-10-027-632-321713

Query Match          85.6%; Score 15.4; DB 13; Length 592;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
; SEQ ID NO: US-10-027-632-321713

Qy      1 GACAAGGCAGGACAGA 17
        ||| | | | | | | | | |
Db      244 GACARGGAGGAGCA 260

Search completed: August 27, 2003, 21:14:03
Job time : 92.7308 secs

RESULT 15
US-10-027-632-286257
; Sequence 286257, Application US/10027632
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-13
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO: 286257
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Human
;   US-10-027-632-286257

Qy      1 GACAAGGCAGGACAGA 17
        ||| | | | | | | | | |
Db      303 GAAAGGGAGGAGCA 287

RESULT 15
US-10-027-632-321713

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Om nucleic - nucleic search, using sw model  
Run on: August 27, 2003, 18:00:44 ; Search time 1013.46 seconds  
          (1372.452 Million cell updates/sec)

Perfect score: 34  
Sequence: US-10-085-612-2  
          1 cgattttgtactgtcaatgtggatcgatggcccc 34

Scoring table: IDENTITY\_NDC  
GapOp 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEnsembl:  
1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_dx:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_yi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_lov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rdl:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

	Result No.	Score	Query Match Length	%	DB ID	Description
	1	34	100.0	830	6	AX472122 Sequence
	2	34	100.0	1612	9	AF325929 Homo sapi
	3	34	100.0	123778	9	AC069294 Homo sapi
	4	33	97.1	442	9	AF315323 Homo sapi
	5	33	97.1	830	6	AX472123 Sequence
	6	33	97.1	830	6	AX472124 Sequence
	7	33	97.1	830	6	AX472125 Sequence
	8	33	97.1	830	6	AX472217 Sequence
	9	33	97.1	830	6	AX472218 Sequence
	10	33	97.1	96960	6	AX706964 Sequence
	11	33	97.1	96960	6	AX707894 Sequence
	12	33	97.1	174832	9	AF280107 Homo sapi
	13	31.4	92.4	830	6	AX472125 Sequence
	14	31.4	92.4	830	6	AX472244 Sequence
	15	30.8	90.6	174004	2	AC141417 Papio anu
	16	27.6	81.2	621	6	AX472134 Sequence
	17	27.6	81.2	624	6	AX472132 Sequence
	18	27.6	81.2	624	6	AX472133 Sequence
	19	27.6	81.2	624	6	AX472235 Sequence
	20	27.6	81.2	624	6	AX472236 Sequence
	21	27.6	81.2	177531	6	AX706962 Sequence
	22	27.6	81.2	177531	6	AX707892 Sequence
	23	27.6	81.2	177531	9	AC0505020 Homo sapi
	c	24	23	67.6	110000	2
	c	25	23	67.6	157888	2
	c	26	23	67.6	171724	2
	c	27	23	67.6	251207	2
	c	28	23	67.6	255495	2
	c	29	23	67.6	271429	2
	c	30	23	67.6	280198	2
	c	31	22.8	67.1	1012	9
	c	32	22.8	67.1	1013	9
	c	33	22.8	67.1	1345	6
	c	34	22.8	67.1	1345	6
	c	35	22.8	67.1	1345	6
	c	36	22.8	67.1	1345	6
	c	37	22.8	67.1	1345	9
	c	38	22.8	67.1	11186	9
	c	39	22.8	67.1	11374	9
	c	40	22.4	65.9	625	11
	c	41	22.4	65.9	89212	9
	c	42	22.4	65.9	138302	9
	c	43	22.4	65.9	154125	2
	c	44	22.4	65.9	168387	9
	c	45	22.4	65.9	177434	2

## ALIGNMENTS

RESULT	1	LOCUS	Sequence 113 from Patent WO20053775.	DNA	linear	PAT 09-AUG-2002
		DEFINITION	AX472122			
		ACCESSION	AX472122			
		VERSION	GI:22207163			
		KEYWORDS				
		SOURCE	Homo sapiens (human)			
		ORGANISM	Homo sapiens			
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;			
			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1	AUTHORS	Hustert, E., Haberl, M. and Wojnowski, L.			
TITLE			Identification of the genetic determinants of the polymorphic cyp3a5 expression			

Pred. No. is the number of results predicted by chance to have a

JOURNAL	Patent: WO 02053775-A 113 11-JUL-2002;	
FEATURES	EPIDAUROS BIOTECHNOLOGIE AG (DE)	
source	Location/Qualifiers	
	1. .830	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
BASE COUNT	229	a 176 c 239 g 186 t
ORIGIN		
RESULT		
AF325929		
LOCUS	AF325929	1612 bp DNA linear
DEFINITION	Homo sapiens cytochrome P450 (CYP3A1) pseudogene, CYP3A1*1 allele, partial sequence, including promoter region.	
ACCESSION	AF325929	PRI 06-DEC-2002
VERSION	AF325929.1	CYP3A1*1
SOURCE		
ORGANISM	Homo sapiens	
MATERIAL	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 ('bases 1 to 1612)	
AUTHORS	Kuehl,P.M., Zhang,J., Lin,Y., Watkins,P., Maurel,P., Daly,A., Wrighton,S., Hall,S., Reiling,M., Schuetz,J., Brimer,C., Storm,S., Bellings,M., Brimer,C., Yasuda,K., Venkataraman,R., Storm,S., Thummel,K., Boguski,M.S. and Schuetz,E.	
TITLE	Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression	
JOURNAL	Nat. Genet. 27 (4), 383-391 (2001)	
MEDLINE	21175742	
PUBMED	11275519	
REFID	2 ('bases 1 to 1612)	
AUTHORS	Kuehl,P.M., Zhang,J., Lin,Y., Watkins,P., Maurel,P., Daly,A., Wrighton,S., Hall,S., Reiling,M., Schuetz,J., Brimer,C., Storm,S., Thummel,K., Boguski,M.S. and Schuetz,E.	
TITLE	Direct Submission (05-DEC-2000) University of Maryland at Baltimore, Baltimore, MD 21201 USA	
JOURNAL	Location/Qualifiers	
FEATURES		
source	1. .1612	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	1. .>1612	
	/gene="CYP3A1"	
gene	/pseudo	
	/allele="CYP3A1*1"	
promoter	1. .1436	
	/gene="CYP3A1"	
	/pseudo	
	1374. .1378	
variation	/note="2 base mismatch and 2 base insertion compared to Genbank Accession Number S74700"	
	/evidence=experimental	
CDS	1540. .>1612	
	/gene="CYP3A1"	
	/note="cytochrome P450"	
	/pseudo	
BASE COUNT	494	a 334 c 388 g 396 t
ORIGIN		
Query Match		

Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
<http://www.ncbi.nlm.nih.gov/DIR/GTB/CHR7/>, send  
 mailto:[egreen@nhgr.nih.gov](mailto:egreen@nhgr.nih.gov), or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhou, B., Frengen, E., Tateno, M., Cataneo, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: PBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is GS1-259H13, 200 bp overlap; the clone sequenced to the left is CTD-3244018, 200 bp overlap. Actual start of this clone is at base position 59660 of CTD-3244018; actual end is at base position 25744 of GS1-259H13.

The sequence from base position 113692 to 114784 is derived from a single plasmid subclone. Assembly in this region is supported by HindIII digest information.

## Location/Qualifiers

1..123778

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RPCI-11-757A13"

/clone\_1="RPCI-11"

repeat\_region

1..337

/rpt\_family="MIR"

repeat\_region

353..657

/rpt\_family="Alu"

repeat\_region

814..928

/rpt\_family="L2"

repeat\_region

1684..1800

/rpt\_family="Alu"

repeat\_region

2040..2334

/rpt\_family="Alu"

repeat\_region

2453..2756

/rpt\_family="Alu"

repeat\_region

3874..4033

/rpt\_family="MER1\_type"

misc\_feature

10188..10513

/note="similar to EST A1377273 (NID:94187126) te65b04.x1"

misc\_feature

10244..10541

/note="similar to EST BF086380 (NID:910892090)"

misc\_feature

10245..10562

/note="similar to EST A977693 (NID:98163072)"

misc\_feature

10255..10556

/note="similar to EST BE218528 (NID:98905846) hv40h11.x1"

misc\_feature

10257..10565

/note="similar to EST BF093582 (NID:910899292)"

misc\_feature

10319..10568

/note="similar to EST AA934962 (NID:93092174) op312g05.s1"

repeat\_region

10569..10682

/rpt\_family="Alu"

repeat\_region

10776..10934

/rpt\_family="Alu"

misc\_feature

10936..11009

/note="similar to EST BE218528 (NID:98905846) hv40h11.x1"

misc\_feature

10942..11021

/note="similar to EST A977693 (NID:98163072)"

repeat\_region

11179..11348

/rpt\_family="ERVI"

11537..11822

/rpt\_family="ERVI"

11814..12038

```

/repeat_region          /note="similar to EST A9890805 (NID:98055010)" 12025..12094
/repeat_region          /rpt_family="ERVI"
/repeat_region          /rpt_family="ERVI" 12095..12505
/repeat_region          /rpt_family="ERVI" 12509..14044
/repeat_region          /rpt_family="Alu" 14045..14130
/repeat_region          /rpt_family="Alu" 14137..14361
/repeat_region          /rpt_family="L1" 14534..15351
/misc_feature           /note="similar to EST A1140604 (NID:g3648061) qe05f02.x1" 15354..15334
/misc_feature           /note="similar to EST AA811274 (NID:g3880885) ob68e03.s1" 15357..15634
/misc_feature           /note="similar to EST AA858281 (NID:g2946583) ob13a02.s1" 15358..15334
/misc_feature           /note="similar to EST AA935054 (NID:g309161) oo68d11.s1" 15373..15634
/misc_feature           /note="similar to EST BE632487 (NID:g10063399)" 15623..1943
/repeat_region          /rpt_family="L1" 19846..20329
/repeat_region          /rpt_family="L1" 20331..20499
/repeat_region          /rpt_family="L1" 20499..20529
/repeat_region          /rpt_family="L1" 20531..20699
/repeat_region          /rpt_family="L1" 20699..21923
/repeat_region          /rpt_family="L1" 22194..23073
/repeat_region          /rpt_family="L1" 23074..23663
/misc_feature           /note="similar to EST AA570297 (NID:g2344277) nm07h03.s1" 23524..23550
/repeat_region          /note="similar to EST AA570297 (NID:g2344277) nm07h03.s1" 24461..24717
/repeat_region          /rpt_family="MIR" 24420..25131
/repeat_region          /rpt_family="Alu" 25223..25918
/repeat_region          /rpt_family="L1" 26005..26043
/repeat_region          /rpt_family="L1" 26102..26483
/repeat_region          /rpt_family="L1" 26485..26797
/repeat_region          /rpt_family="Alu" 27787..27913
/repeat_region          /rpt_family="L1" 27936..28269
/repeat_region          /rpt_family="L1" 28222..28334
/repeat_region          /rpt_family="M4LR" 28384..28690
/repeat_region          /rpt_family="Alu" 28815..28939
/repeat_region          /rpt_family="L2" 29027..29294
/repeat_region          /rpt_family="L2" 30023..30152
/repeat_region          /rpt_family="M4LR" 3030..30377
/repeat_region          /rpt_family="MIR" 30356..30556
Query Match           100.0% Score 34; DB 9; Length 123778;
Best Local Similarity 100.0% Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 CGATTCTTGTACTGGTGAGCTGAGCCCCG 34
Db      119173 CGATTCTTGTACTGGTGAGCCCCG 119206

```

RESULT 4  
 AF31532054 AF31532054 442 bp DNA linear PRI 07-FEB-2001  
 LOCUS Homo sapiens CYP3A2P pseudogene, partial sequence.  
 DEFINITION AF315323  
 ACCESSION AF315323.1 GI:12082805  
 VERSION  
 KEYWORDS  
 SEGMENT 4 of 5  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 442)  
 AUTHORS Finta, C. and Zaphiroopoulos P.G.  
 TITLE The human cytochrome P450 3A locus. Gene evolution by capture of downstream exons.  
 JOURNAL Gene 260 (1-2), 13-23 (2000)  
 MEDLINE 20578884  
 PUBMED 11137287  
 REFERENCE 2 (bases 1 to 442)  
 AUTHORS Finta, C. and Zaphiroopoulos P.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-OCT-2000) Department of Biosciences, Karolinska Institute, NOVUM, Ruddinge 14157, Sweden  
 FEATURES Location/Qualifiers  
 SOURCE 1. -442  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q21-q22.1"  
 /clone="CITB BAC 128D24"  
 /pseudo="CITB BAC 128D24"  
 misc\_feature 246..316  
 /gene="CYP3A2P2"  
 /note="exon 1-like sequence; mono-oxygenase"  
 BASE COUNT 112 a 106 c 111 g 113 t  
 ORIGIN  
 Query Match 2 GATCTTGCTACTGGCTCAGCTGAGCCCC 34  
 Best Local Similarity 97.1%; Score 33; DB 9; Length 442;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 67 GATTCCTTGTACTGGCTCAGCTGAGCCCC 99  
 Db 67 GATTCCTTGTACTGGCTCAGCTGAGCCCC 99  
 RESULT 5  
 AX472123 AX472123 830 bp DNA linear PAT 09-AUG-2002  
 LOCUS Sequence 114 from Patent WO02053775.  
 DEFINITION AF472123  
 ACCESSION AX472123.1 GI:22207164  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 Hustert, E., Haberl, M. and Wojnowski, L.  
 AUTHORS  
 TITLE Identification of the genetic determinants of the polymorphic CYP3A5 expression  
 JOURNAL Patent: WO 02053775-A 114 11-JUL-2002;  
 EPIDAUROS BIOTECHNOLOGIE AG (DE)  
 FEATURES Location/Qualifiers  
 SOURCE 1..830  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 231 a 176 c 236 g 187 t  
 ORIGIN  
 Query Match 1 CGATCTTGTACTGGCTCAGCTGAGCCCC 33  
 Best Local Similarity 97.1%; Score 33; DB 6; Length 830;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGATCTTGTACTGGCTGAAGCTGCAAGCCCC 33  
 Db 549 CGATCTTGTACTGGCTGAAGCTGCAAGCCCC 581

RESULT 8  
 AX472217 LOCUS AX472217 Sequence 208 from Patent WO20533775. linear PAT 09-AUG-2002  
 DEFINITION AX472217 VERSION AX472217.1 GI:22207254  
 KEYWORDS Homo sapiens (human)  
 ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 Heinrich,G. and Kerb,R.  
 AUTHORS Methods for the treatment of cancer with irinotecan based on CYP3A5  
 TITLE Patent: WO 03013534 A662 20-FEB-2003;  
 JOURNAL Epidauros Biotechnologie AG (DE)  
 FEATURES Location/Qualifiers  
 SOURCE 1. .96360  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 27407 a 19882 c 20450 g 29221 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 96960;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GATTCTTGCTACTGGCGCAGCTGCAGCCCC 34  
 Db 93439 GATTCTTGCTACTGGCGCAGCTGCAGCCCC 93471

RESULT 11  
 AX470894 LOCUS AX707894 Sequence 662 from Patent WO03013536. linear PAT 04-APR-2003  
 DEFINITION AX707894 VERSION AX707894 GI:29563969  
 KEYWORDS Homo sapiens (human)  
 ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 Heinrich,G. and Kerb,R.  
 AUTHORS Methods for treatment of cancer using irinotecan based on UST1A1  
 TITLE Patent: WO 0301336 A662 20-FEB-2003;  
 JOURNAL Epidauros Biotechnology AG (DE)  
 FEATURES Location/Qualifiers  
 SOURCE 1. .96360  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 27407 a 19882 c 20450 g 29221 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 96960;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GATTCTTGCTACTGGCTGAGCTGCAGCCCC 34  
 Db 93439 GATTCTTGCTACTGGCTGAGCTGCAGCCCC 93471

RESULT 9  
 AX472218 LOCUS AX472218 Sequence 209 from Patent WO2053775. linear PAT 09-AUG-2002  
 DEFINITION AX472218 VERSION AX472218.1 GI:22207255  
 KEYWORDS Homo sapiens (human)  
 ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 Hustert,E., Haberl,M. and Wojnowski,L.  
 AUTHORS Identification of the genetic determinants of the polymorphic  
 TITLE CYP3A5 expression  
 JOURNAL Patent: WO 0305375 A 209 11-JUL-2002;  
 EPIDAUROS BIOTECHNOLOGIE AG (DE)  
 FEATURES Location/Qualifiers  
 SOURCE 1. .830  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 230 a 177 c 237 g 186 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 830;  
 Best Local Similarity 100.0%; Pred. No. 0.0034;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATCTTGTACTGGCTGAAGCTGCAAGCCCC 33  
 Db 549 CGATCTTGTACTGGCTGAAGCTGCAAGCCCC 581

RESULT 10  
 AX706964 LOCUS AX706964 Sequence 662 from Patent WO03013534. linear PAT 04-APR-2003  
 DEFINITION AX706964 VERSION AX706964 GI:29563289  
 KEYWORDS Homo sapiens (human)  
 ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 Heinrich,G. and Kerb,R.  
 AUTHORS Methods for the treatment of cancer with irinotecan based on CYP3A5  
 TITLE Patent: WO 03013534 A662 20-FEB-2003;  
 JOURNAL Epidauros Biotechnologie AG (DE)  
 FEATURES Location/Qualifiers  
 SOURCE 1. .96360  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 27407 a 19882 c 20450 g 29221 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 96960;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GATTCTTGCTACTGGCTGAGCTGCAGCCCC 34  
 Db 93439 GATTCTTGCTACTGGCTGAGCTGCAGCCCC 93471

RESULT 12  
 AF280107 LOCUS AF280107 Homo sapiens cytochrome P450 polypeptide 43 (CYP4A3) gene, partial  
 DEFINITION Homo sapiens cytochrome P450 polypeptide 4 (CYP4A4) and cytochrome P450  
 genes, complete cds; and cytochrome P450  
 polypeptide 7 (CYP3A7) genes, complete cds; and cytochrome P450  
 polypeptide 5 (CYP3A5) gene, partial cds





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 17:59:19 ; Search time 264.808 Seconds  
(without alignments)  
346.595 Million cell updates/sec

Title: US-10-085-612-2

Perfect score: 34  
Sequence: 1 cgatttttgcgttgcggctcgagcccg 34

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs. 1349719017 residues

Total number of hits satisfying chosen parameters:

5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03 : \*

- 1: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDS1/gcadata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	34	100.0	34	AAD45761	Human cytochrome P450 gene
2	34	100.0	830	ABK99409	Human CYP3A5 gene
3	34	100.0	123785	ABX7171	DNA sequence of hu
4	33	97.1	34	AAD45763	Human cytochrome P
5	33	97.1	830	ABK99410	Human CYP3A5 gene
6	33	97.1	830	ABK99411	Human CYP3A5 gene
7	33	97.1	830	ABK99499	Human CYP3A5 gene
8	33	97.1	830	ABK9500	Human CYP3A5 gene

Homo sapiens.

OS

Synthetic.

OS

Key

FT variation

Location/Qualifiers

/standard\_name= "Single nucleotide polymorphism (SNP)"

/tag= a

/replace (34, A)

WO20026448-A1.

PN

06-SEP-2002.

PP

26-FEB-2002; 2002WO-US06115.

PR

26-FEB-2001; 2001US-271630P.

XX

XX&lt;/

PA (DNAS-) DNA SCI INC.  
 PA (UYD-) UNTV DUKE.  
 XX Guida M, Hall J, Petros WP, Vredenburgh JJ, Colvin OM, Marks JR;  
 PI XX Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
 PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
 PT polymorphisms -  
 XX DR; 2002-691652/74.  
 XX PT New nucleic acid molecule useful for identifying polymorphisms  
 PT associated with CYP3A4, CYP3A5 or GSTM1 substrates which influence  
 PT breast cancer treatments, comprises at least one base variation from  
 PT human CYP3A4 or CYP3A5 sequence -  
 XX PS; Claim 9; Page 12; 41pp; English.  
 XX The invention relates to a nucleic acid molecule comprising at least one  
 CC base variation from human cytochrome P450 (CYP) 3A4 or CYP3A5 sequence.  
 CC Nucleic acid molecules of the invention are useful for identifying  
 CC polymorphisms associated with CYP3A4, CYP3A5 or glutathione S-transferase  
 CC (GST) M1 substrates which influence breast cancer treatments. They are  
 CC also useful in diagnostic purposes to identify individuals having a  
 CC polymorphic genotype which influence the outcome of breast cancer  
 CC treatments and the selection of chemotherapeutic agents used to treat  
 CC breast cancer. The polymorphisms depicted are used to screen altered  
 CC metabolism of CYP3A4, CYP3A5 or GSTM1 substrates, potential drug-drug  
 CC interactions, drug adverse effects, likelihood of successful clinical  
 CC outcome following treatment with anti-cancer agents such as cisplatin,  
 CC cyclophosphamide and/or BCNU. The present sequence is human CYP3A5  
 CC polymorphic variant DNA fragment.  
 XX Sequence 34 BP; 4 A; 12 C; 9 G; 9 T; 0 other;  
 SQ Query Match 1 CGATCTTGTGCTACTGGCTCAGTCAGCCCC 34  
 Best Local Similarity 100.0%; Score 34; DB 24; Length 34;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db RESULT 3  
 ABX77171 standard; DNA; 123785 BP.  
 XX AC ABX77171;  
 XX DT 25-APR-2003 (first entry)  
 ID XX DE DNA sequence of human BAC clone RP1-757A13.  
 AC XX KW Human; ss; transgenic; drug metabolism; behaviour; gene; mouse;  
 KW DE pharmacokinetic assay; pharmacodynamic assay; toxicology; serum albumin;  
 KW XX KW alpha-acidic glycoprotein; CYP; multidrug resistance protein; MRP;  
 KW XX KW uridine diphosphoglucuronosyl transferase; UGT; cytochrome P450.  
 OS Homo sapiens.  
 XX PN WO200283897-A1.  
 XX PD 24-OCT-2002.  
 XX PF 18-APR-2002; 2002WO-AU00485.  
 XX PR 18-APR-2001; 2001AU-0004467.  
 XX PA (GENE-) GENE STREAM PTY LTD.  
 XX PI Daly JM;  
 XX DR WPI; 2003-093021/08.  
 XX PR 29-DEC-2000; 2000EP-0128627.  
 XX PR 28-DEC-2000; 2000EP-0128627.  
 XX PR 28-DEC-2000; 2000EP-0128627.  
 XX PR 29-DEC-2000; 2000EP-0128627.  
 XX PR 16-JAN-2001; 2001EP-010012.  
 XX PR 18-JAN-2001; 2001EP-0128627.  
 XX PR 15-AUG-2001; 2001EP-0118884.  
 XX PR 16-AUG-2001; 2001US-312823P.  
 XX PA (EPIDAUROS BIOTECHNOLOGIE AG.  
 XX PI Wojnowski L, Haberl M, Hustert E;  
 XX PS Example 2A; Page 229-295; 408pp; English.  
 XX CC This invention relates to a transgenic non-human animal which may be

XX DR WPI; 2002-583628/62.  
 XX PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
 PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
 PT polymorphisms -  
 XX PS; Claim 1; Figure 4; 138pp; English.  
 XX The present invention relates to a new CYP3A5 polynucleotide encoding a  
 CC polypeptide, where the polynucleotide is capable of hybridising to a  
 CC CYP3A5 gene. The invention is useful in an in vitro method for  
 CC identifying a polymorphism. The invention is also useful for useful for  
 CC diagnosing a disorder related to the presence of a molecular variant of a  
 CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
 CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
 CC The invention can further be used for the preparation of a diagnostic  
 CC composition for diagnosing a disease in a subject having a genome  
 CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
 CC African American. The molecules of the invention are as forensic markers  
 CC and in pharmacological studies. The present nucleic acid sequence  
 CC represents a genomic DNA sequence that contains a region in which a  
 CC polymorphism of the human CYP3A5 gene was detected.  
 XX Sequence 830 BP; 229 A; 176 C; 239 G; 186 T; 0 other;  
 SQ Query Match 1 CGATCTTGTGCTACTGGCTCAGTCAGCCCC 34  
 Best Local Similarity 100.0%; Pred. No. 0.0012;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 549 CGATCTTGTGCTACTGGCTCAGTCAGCCCC 582  
 RESULT 3  
 ABX77171 standard; DNA; 123785 BP.  
 XX AC ABX77171;  
 XX DT 25-APR-2003 (first entry)  
 ID XX DE DNA sequence of human BAC clone RP1-757A13.  
 AC XX KW Human; ss; transgenic; drug metabolism; behaviour; gene; mouse;  
 KW DE pharmacokinetic assay; pharmacodynamic assay; toxicology; serum albumin;  
 KW XX KW alpha-acidic glycoprotein; CYP; multidrug resistance protein; MRP;  
 KW XX KW uridine diphosphoglucuronosyl transferase; UGT; cytochrome P450.  
 OS Homo sapiens.  
 XX PN WO200283897-A1.  
 XX PD 24-OCT-2002.  
 XX PF 18-APR-2002; 2002WO-AU00485.  
 XX PR 18-APR-2001; 2001AU-0004467.  
 XX PA (GENE-) GENE STREAM PTY LTD.  
 XX PI Daly JM;  
 XX DR WPI; 2003-093021/08.  
 XX PT New transgenic non-human animal expressing a foreign polypeptide  
 PT associated with drug behavior and/or metabolism, useful for studying  
 PT the behavior and/or metabolism of a drug in other animals -  
 XX PS Example 2A; Page 229-295; 408pp; English.  
 XX CC This invention relates to a transgenic non-human animal which may be



comprising a variant allele of the CYP3A5 gene, where the subject is an African American. The molecules of the invention are as forensic markers and in Pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human CYP3A5 gene was detected.

Sequence 830 BP; 231 A; 176 C; 236 G; 187 T; 0 other;

Query Match	97.1%	Score	33;	DB	24;	Length	830;
Best Local Similarity	100.0%	Pred. No.	0.0029;				
Matches	33;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps							

RESULT 6

KX99411 ABK9411 standard; DNA; 830 BP.

ABK9411;

21-OCT-2002 (first entry)

Human CYP3A5 gene polymorphic DNA sequence #3.

Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes; AIDS; African American; forensic marker; pharmacological; cytostatic; antidiabetic; anti-HIV; gene therapy; ds.

Homo sapiens.

WO200253775-A2.

11-JUL-2002.

21-DEC-2001; 2001WO-EP15290.

28-DEC-2000; 2000EP-0128627.

28-DEC-2000; 2000US-258634P.

29-DEC-2000; 2000US-258932P.

16-JAN-2001; 2001EP-0100172.

18-JAN-2001; 2001US-262839P.

16-AUG-2001; 2001EP-0118884.

16-AUG-2001; 2001US-312835P.

(EPTD-) EPTDAUROS BIOTECHNOLOGIE AG.

Wojnowski L, Habert M, Hustert E;

WPI: 2002-583628/62.

Novel CYP3A5 polynucleotide useful for diagnosis and treatment of cancer, cardiovascular diseases, diabetes and AIDS, and for identifying polymorphisms -

Claim 1: Figure 4; 138pp; English.

The present invention relates to a new CYP3A5 polynucleotide encoding a polypeptide, where the polynucleotide is capable of hybridising to a CYP3A5 gene. The invention is useful in an in vitro method for identifying a polymorphism. The invention is also useful for diagnosing a disorder related to the presence of a molecular variant of CYP3A5 or susceptibility to such a disorder, where the disorder is cancer, or diseases including cardiovascular diseases, diabetes and AIDS. The invention can further be used for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the cyp3A5 gene, where the subject is an African American. The molecules of the invention are as forensic markers and in pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human CYP3A5 gene was detected.

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9  
 ID ABK9500 standard; DNA; 830 BP.  
 XX  
 AC ABK9500;  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Human CYP3A5 gene polymorphic DNA sequence # 31.  
 XX Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
 KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200253775-A2.  
 PN PD 11-JUL-2002.  
 XX PF 21-DEC-2001; 2001WO-EP15290.  
 PR 28-DEC-2000; 2000EP-0128627.  
 PR 28-DEC-2000; 2000TS-258684P.  
 PR 29-DEC-2000; 2000US-25852P.  
 PR 29-DEC-2000; 2000US-25854P.  
 PR 16-FAN-2001; 2001EP-010172.  
 PR 18-FAN-2001; 2001US-26259P.  
 PR 16-AUG-2001; 2001EP-011884.  
 PR 16-AUG-2001; 2001US-31283P.  
 PA (EPID-) EPIDAURUS BIOTECHNOLOGIE AG.  
 XX PI Wojnowski L, Haberl M, Hustert E;  
 XX DR WPI; 2002-583628/62.  
 XX PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
 PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
 PT polymorphisms -  
 XX PS Claim 1; Figure 4; 138pp; English.  
 XX CC The present invention relates to a new CYP3A5 polynucleotide encoding a  
 CC polypeptide, where the polynucleotide is capable of hybridising to a  
 CC CYP3A5 gene. The invention is useful in an in vitro method for  
 CC identifying a polymorphism. The invention is also useful for useful for  
 CC diagnosing a disorder related to the presence of a molecular variant of a  
 CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
 CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
 CC The invention can further be used for the preparation of a diagnostic  
 CC composition for diagnosing a disease in a subject having a genome  
 CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
 CC African American. The molecules of the invention are as forensic markers  
 CC and in pharmacological studies. The present nucleic acid sequence  
 CC represents a genomic DNA sequence that contains a region in which a  
 CC polymorphism of the human CYP3A5 gene was detected.  
 XX SQ Sequence 830 BP; 230 A; 175 C; 237 G; 188 T; 0 other;  
 Query Match 97.1%; Score 33; DB 24; Length 830;  
 Best Local Similarity 100.0%; Pred. No. 0.0039;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 10  
 AAA51756  
 ID AAA51756 standard; DNA; 1346 BP.

RESULT 8  
 ID ABK9500 standard; DNA; 830 BP.  
 XX  
 AC ABK9501;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX Human CYP3A5 gene polymorphic DNA sequence # 32.  
 XX Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
 KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200253775-A2.  
 PN PD 11-JUL-2002.  
 XX PF 21-DEC-2001; 2001WO-EP15290.  
 PR 28-DEC-2000; 2000EP-0128627.  
 PR 28-DEC-2000; 2000TS-258684P.  
 PR 29-DEC-2000; 2000US-25895P.  
 PR 16-JAN-2001; 2001EP-0100172.  
 PR 18-JAN-2001; 2001US-36285P.  
 PR 16-AUG-2001; 2001EP-011884.  
 PR 16-AUG-2001; 2001US-31282P.  
 PA (EPID-) EPIDAURUS BIOTECHNOLOGIE AG.  
 XX PI Wojnowski L, Haberl M, Hustert E;  
 XX DR WPI; 2002-583628/62.  
 XX PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
 PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
 PT polymorphisms -  
 XX PS Claim 1; Figure 4; 138pp; English.  
 XX CC The present invention relates to a new CYP3A5 polynucleotide encoding a  
 CC polypeptide, where the polynucleotide is capable of hybridising to a  
 CC CYP3A5 gene. The invention is useful in an in vitro method for  
 CC identifying a polymorphism. The invention is also useful for useful for  
 CC diagnosing a disorder related to the presence of a molecular variant of a  
 CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
 CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
 CC The invention can further be used for the preparation of a diagnostic  
 CC composition for diagnosing a disease in a subject having a genome  
 CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
 CC African American. The molecules of the invention are as forensic markers  
 CC and in pharmacological studies. The present nucleic acid sequence  
 CC represents a genomic DNA sequence that contains a region in which a  
 CC polymorphism of the human CYP3A5 gene was detected.  
 XX SQ Sequence 830 BP; 230 A; 177 C; 237 G; 186 T; 0 other;  
 Query Match 97.1%; Score 33; DB 24; Length 830;  
 Best Local Similarity 100.0%; Pred. No. 0.0029;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 10  
 AAA51756  
 ID AAA51756 standard; DNA; 1346 BP.

XX AAA51756;  
 XX AC  
 XX DT 31-OCT-2000 (first entry)  
 DE Cytochrome P450 CYP3A5 gene 5' flanking region (-1343 to +3).  
 XX CYP3A5; Cytochrome P450; transcription regulatory region; polymorphism;  
 KW Activator protein-3 motif; Ap-3; basic transcription element;  
 KW drug metabolism; phenotype; ss.  
 XX OS Homo sapiens.  
 XX Key  
 FH variation  
 FT replace(863,G)  
 FT /\*tag= a  
 /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "Corresponds to nucleotide position -45. Presence  
 of a G in variants destroys an Ap-3 site."  
 FT variation  
 FT replace(1197,G)  
 FT /\*tag= b  
 /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "Corresponds to nucleotide position -147. Presence  
 of a G in variants destroys a basic transcription  
 element."  
 XX WO200039332-A1.  
 XX 06-JUL-2000.  
 XX 22-DEC-1999; 99NO-GB04380.  
 XX 23-DEC-1998; 98GB-0028619.  
 XX (JANC ) JANSSEN PHARM NV.  
 XX Paulussen ADC, Armstrong M;  
 PI WTI; 2000-452418/39.  
 DR PS  
 XX PT Identifying subjects with a high drug metabolizing phenotype associated  
 PT with cytochrome CYP3A5 expression for establishing whether a drug will  
 PT be metabolized by the subject  
 XX PS  
 XX Claim 5; Fig 7; 68pp; English.  
 CC Cytochrome P450 subfamily CYP3A5 transcription regulatory regions can be  
 CC screened for the presence/absence of a polymorphic variant, preferably  
 CC at positions -475 or -147 of the 5' flanking region adjacent  
 CC to the CYP3A5 coding sequence. The variants are present in an activator  
 CC protein-3 (Ap-3) motif and/or a basic transcription element (BTE). The  
 CC polymorphisms cause increased CYP3A5 gene expression and this has been  
 CC linked to drug metabolic activity. Screening for the presence of variants  
 CC can be used to identify subjects with a high or low drug metabolizing  
 CC phenotype associated with cytochrome CYP3A5 expression. Primers are used  
 CC in addition to hybridizing to the site of interest, are capable of  
 CC introducing a restriction site which is absent in either the wild type  
 CC sequence or polymorphic variants. Restriction enzyme cleavage analysis  
 CC can then be used to establish, before treatment with a drug, whether the  
 CC methods are used to effectively metabolized by the patient, to identify  
 CC compounds and transcription factors that can bind to a DNA sequence  
 CC encoding CYP3A5, diagnosing susceptibility to a disease which is caused  
 CC by toxins or procarcinogens metabolized by CYP3A5 and for identifying  
 CC mutagenic effects of a compound.  
 XX Sequence 1346 BP; 409 A; 275 C; 344 G; 318 T; 0 other;  
 SQ Query Match 97.1%; Score 33; DB 21; Length 1346;  
 Best Local Similarity 100.0%; Pred. No. 0.0031;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 549 CAATCTTGTGACTGGCTACGGTCAGTCAGTGAGCCCC 33  
 RESULT 12

XX Db 1164 CGATTCTTGTGACTGGCTACGGCTAGTCAGTGAGCCCC 1196  
 RESULT 11  
 ID ABK9412 standard; DNA; 830 BP.  
 XX ID ABK9412;  
 AC ABK9412;  
 XX DT 21-OCT-2002 (first entry)  
 XX Human CYP3A5 gene polymorphic DNA sequence #4.  
 XX Human; CYP3A5; Polymorphism; cancer; cardiovascular disease; diabetes;  
 KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX DE Human CYP3A5 gene polymorphic DNA sequence #4.  
 XX Human CYP3A5; Polymorphism; cancer; cardiovascular disease; diabetes;  
 KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX OS Homo sapiens.  
 XX PN WO200253775-A2.  
 XX PD 11-JUL-2002.  
 PF 21-DEC-2001; 2001WO-EP15290.  
 XX PR 28-DEC-2000; 2000EP-01286627.  
 PR 28-DEC-2000; 2000US-258848P.  
 PR 29-DEC-2000; 2000US-258932P.  
 PR 16-JAN-2001; 2001EP-0100172.  
 PR 18-JAN-2001; 2001US-262859P.  
 PR 16-AUG-2001; 2001EP-0118884.  
 PR 16-AUG-2001; 2001US-312823P.  
 XX PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.  
 PA PI Wojnowski L, Haberl M, Hustert E;  
 XX XX DR WPI; 2002-583628/62.  
 XX PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
 PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
 PT polymorphisms -  
 XX PS  
 XX Claim 1; Figure 4; 138pp; English.  
 CC The present invention relates to a new CYP3A5 polynucleotide encoding a  
 CC polypeptide, where the polynucleotide is capable of hybridising to a  
 CC CYP3A5 gene. The invention is useful in an in vitro method for  
 CC identifying a polymorphism. The invention is also useful for useful for  
 CC diagnosing a disorder related to the presence of a molecular variant of a  
 CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
 CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
 CC The invention can further be used for the preparation of a diagnostic  
 CC composition for diagnosing a disease in a subject having a genome  
 CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
 CC African American. The molecules of the invention are as forensic markers  
 CC and in pharmacological studies. The present nucleic acid sequence  
 CC represents a genomic DNA sequence that contains a region in which a  
 CC polymorphism of the human CYP3A5 gene was detected.  
 XX SQ Sequence 830 BP; 231 A; 176 C; 236 G; 187 T; 0 other;  
 Query Match 92.4%; Score 31.4; DB 24; Length 830;  
 Best Local Similarity 97.0%; Pred. No. 0.012; Mismatches 0; Indels 0; Gaps 0;  
 Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGATCTTGTGACTGGCTACGGTCAGTCAGTGAGCCCC 33  
 Db 549 CAATCTTGTGACTGGCTACGGTCAGTCAGTGAGCCCC 581  
 RESULT 12

QY 1 CGATCTTGTGACTGGCTACGGTCAGTCAGTGAGCCCC 33

ABK9524	ID ABK9524 standard; DNA; 830 BP.	DT 21-OCT-2002 (first entry)
XX	XX	XX
AC	AC	DE Human CYP3A5 gene polymorphic DNA sequence #13.
XX	XX	XX
21-OCT-2002	(first entry)	KW Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes; AIDS; African American; forensic marker; pharmacological; cytostatic; antidiabetic; anti-HIV; gene therapy; ds.
XX	XX	KW antidiabetic; anti-HIV; gene therapy; ds.
DE Human CYP3A5 gene polymorphic DNA sequence #55.	XX	XX
XX	OS Homo sapiens.	XX
XX	XX	XX
PN WO200253775-A2.	PN	XX
XX	XX	XX
PD 11-JUL-2002.	PD	XX
XX	XX	XX
21-DEC-2001;	2001WO-EP15290.	XX
XX	XX	XX
PR 28-DEC-2000;	2000EP-0123627.	XX
PR 28-DEC-2000;	2000US-258384P.	PR
PR 29-DEC-2000;	2000US-258352P.	PR
PR 16-JAN-2001;	2001EP-0100172.	PR
PR 18-JAN-2001;	2001US-26259P.	PR
PR 16-AUG-2001;	2001EP-0118894.	PR
PR 16-AUG-2001;	2001US-312825P.	PR
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.	XX	XX
XX	XX	XX
PI Wojnowski L, Haberl M, Hustert E;	PI	XX
XX	XX	XX
DR WPI; 2002-583628/62.	DR	XX
XX	XX	XX
PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of cancer, cardiovascular diseases, diabetes and AIDS, and for identifying polymorphisms -	PT	XX
XX	XX	XX
PS Claim 1; Figure 4; 138pp; English.	PS	XX
XX	XX	XX
PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of cancer, cardiovascular diseases, diabetes and AIDS, and for identifying polymorphisms -	PT	CC
XX	XX	CC
DR WPI; 2002-583628/62.	DR	CC
XX	XX	CC
PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of cancer, cardiovascular diseases, diabetes and AIDS, and for identifying polymorphisms -	PT	CC
XX	XX	CC
PS Claim 1; Figure 4; 138pp; English.	PS	CC
XX	XX	CC
CC The present invention relates to a new CYP3A5 polynucleotide encoding a polypeptide, where the polynucleotide is capable of hybridising to a CYP3A5 gene. The invention is useful in an in vitro method for identifying a polymorphism. The invention is also useful for useful for diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 or susceptibility to such a disorder, where the disorder is cancer, or diseases including cardiovascular diseases, diabetes and AIDS. The invention can further be used for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the CYP3A5 gene, where the subject is an African American. The molecules of the invention are as forensic markers and in pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human CYP3A5 gene was detected.	CC	
XX	XX	CC
CC The present invention relates to a new CYP3A5 polynucleotide encoding a polypeptide, where the polynucleotide is capable of hybridising to a CYP3A5 gene. The invention is useful in an in vitro method for identifying a polymorphism. The invention is also useful for useful for diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 or susceptibility to such a disorder, where the disorder is cancer, or diseases including cardiovascular diseases, diabetes and AIDS. The invention can further be used for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the CYP3A5 gene, where the subject is an African American. The molecules of the invention are as forensic markers and in pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human CYP3A5 gene was detected.	CC	
XX	XX	CC
PS Sequence 830 BP; 230 A; 177 C; 237 G; 186 T; 0 other;	PS	CC
XX	XX	CC
Query Match Score 92.4%; Best Local Similarity 97.0%; Matches 32; Conservative 0; Mismatches 0;	QY	1 CGATTTGGTACTGGCTCAGTGAGCCCC 34
DB 549 CGATTCATTGCTTGTACCGGCTGAGCTGAGCCCC 581	DB	65 CGATTCATTGCTTGTACCGGCTGAGCTGAGCCCC 98
XX	XX	XX
RESULT 13 ABK9419	RESULT 14 ABK9419	XX
ID ABK9421 standard; DNA; 621 BP.	ID ABK9419 standard; DNA; 624 BP.	AC ABK9419;
XX	XX	XX
AC ABK9421;	DE Human CYP3A5 gene polymorphic DNA sequence #11.	KW Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;
XX	XX	XX

KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX  
 XX OS PN WO200253775-A2.  
 Homo sapiens.  
 XX OS PN WO200253775-A2.  
 XX PD 11-JUL-2002.  
 XX PF 21-DEC-2001; 2001WO-EP15290.  
 XX PR 28-DEC-2000; 2000EP-0128627.  
 XX PR 28-DEC-2000; 2000US-251684P.  
 XX PR 29-DEC-2000; 2000US-258152P.  
 XX PR 28-DEC-2000; 2000US-263859P.  
 XX PR 29-DEC-2000; 2000US-258952P.  
 XX PR 16-JAN-2001; 2001EP-0110172.  
 XX PR 18-JAN-2001; 2001US-263859P.  
 XX PR 16-AUG-2001; 2001EP-011884.  
 XX PR 16-AUG-2001; 2001US-312825P.  
 XX PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.  
 XX PA Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
 PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
 PT polymorphisms -  
 XX PS Claim 1; Figure 4; 138pp; English.  
 XX CC The present invention relates to a new CYP3A5 polynucleotide encoding a  
 CC polypeptide, where the polynucleotide is capable of hybridising to a  
 CC CYP3A5 gene. The invention is useful in an in vitro method for  
 CC identifying a polymorphism. The invention is also useful for useful for  
 CC diagnosing a disorder related to the presence of a molecular variant of a  
 CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
 CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
 CC The invention can further be used for the preparation of a diagnostic  
 CC composition for diagnosing a disease in a subject having a genome  
 CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
 CC African American. The molecules of the invention are as forensic markers  
 CC and in pharmacological studies. The present nucleic acid sequence  
 CC represents a genomic DNA sequence that contains a region in which a  
 CC polymorphism of the human CYP3A5 gene was detected.  
 XX Sequence 624 BP; 174 A; 134 C; 149 G; 167 T; 0 other;  
 XX Query Match 81.2%; Score 27.6; DB 24; Length 624;  
 XX Best Local Similarity 88.2%; Pred. No. 0.34;  
 XX Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 XX Sq 1 CGATTCTTGTACTGGCTGAGGTGAGCCCG 34  
 QY 1 CGATTCCTTGCTACTGGCTGAGGTGAGCCCG 34  
 XX 65 CGATTCCTTGCTATGGCTGAGCTATGGCTGAGCTATAGCCCTG 98  
 Db Search completed: August 27, 2003, 18:15:45  
 XX Job time : 265.808 secs  
 XX  
 RESULT 15  
 ID ABR99420  
 XX DE Human CYP3A5 gene polymorphic DNA sequence #12.  
 AC ABR99420;  
 XX DT 21-OCT-2002 (first entry)  
 XX KW Human; CYP3A5; Polymorphism; cancer; cardiovascular disease; diabetes;  
 KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX OS Homo sapiens.  
 XX

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - nucleic search, using sw model

on: August 27, 2003, 18:02:39 ; Search time 37.0385 Seconds  
(without alignments)  
214.504 Million cell updates/sec

hitlist:  
perfect score: 18  
sequence: 1 gacaaggcaggacagaa 18  
scoring table: IDENTITY\_NUC Gapop 10.0 , Gapext 1.0  
searched: 569978 seqs, 220691666 residues

total number of hits satisfying chosen parameters: 1139956

minimum DB seq length: 0  
maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cpn2\_6/podata/1/ina/5A.COMB.seq.\*
- 2: /cpn2\_6/podata/1/ina/5B.COMB.seq.\*
- 3: /cpn2\_6/podata/1/ina/6A.COMB.seq.\*
- 4: /cpn2\_6/podata/1/ina/6B.COMB.seq.\*
- 5: /cpn2\_6/podata/1/ina/PCMV.COMB.seq.\*
- 6: /cpn2\_6/podata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Query Score	Match	Length	DB ID	Description
c 1	16.4	91.1	29	4 US-09-144-367-13	Sequence 13, Appl
c 2	16.4	1345	3	US-09-144-367-13	Sequence 2, Appl
c 3	16.8	88.9	1345	4 US-09-144-367-13	Sequence 3, Appl
c 4	15.4	85.6	19	4 US-09-144-367-14	Sequence 44, Appl
c 5	15.4	85.6	45546	4 US-09-144-053-6	Sequence 6, Appl
c 6	14.8	82.2	32	4 US-09-144-367-12	Sequence 12, Appl
c 7	14.8	82.2	807	1 US-08-154-019-28	Sequence 28, Appl
c 8	14.8	82.2	807	1 US-08-146-333-28	Sequence 28, Appl
c 9	14.8	82.2	807	3 US-08-464-167-28	Sequence 28, Appl
c 10	14.8	82.2	807	3 US-09-158-313-28	Sequence 28, Appl
c 11	14.8	82.2	807	3 US-08-476-798-28	Sequence 28, Appl
c 12	14.8	82.2	824	1 US-08-154-018-29	Sequence 29, Appl
c 13	14.8	82.2	824	1 US-08-154-019-28	Sequence 29, Appl
c 14	14.8	82.2	824	3 US-08-464-167-29	Sequence 29, Appl
c 15	14.8	82.2	824	3 US-09-155-313-29	Sequence 29, Appl
c 16	14.8	82.2	824	3 US-08-476-798-29	Sequence 29, Appl
c 17	14.8	82.2	1202	4 US-09-904-615-63	Sequence 63, Appl
c 18	14.8	82.2	1345	3 US-09-372-339-1	Sequence 1, Appl
c 19	14.8	82.2	1452	4 US-09-904-615-34	Sequence 34, Appl
c 20	14.8	82.2	1950	4 US-09-205-258-131	Sequence 131, Appl
c 21	14.8	82.2	4523	4 US-09-473-716-1	Sequence 1, Appl
c 22	14.8	82.2	6763	2 US-08-756-506-23	Sequence 23, Appl
c 23	14.8	82.2	10807	1 US-08-206-176-5	Sequence 7, Appl
c 24	14.8	82.2	10807	2 US-08-756-506-2	Sequence 5, Appl
c 25	14.8	82.2	16063	4 US-09-801-054-3	Sequence 3, Appl
c 26	14.8	82.2	2095	4 US-09-729-954-3	Sequence 17, Appl
c 27	14.8	82.2	48974	3 US-09-920-442-17	Sequence 2, Appl

ALIGNMENTS

RESULT 1	US-0-144-367-13/c	TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
/ Sequence 13, Application US/09144367	/ Patent No. 6431639	/ GENERAL INFORMATION:
/ APPLICANT: Guido, Marco	/ FILE REFERENCE: SEQ-12P	/
/	/ CURRENT APPLICATION NUMBER: US/09/144-367	/
/	/ CURRENT FILING DATE: 1998-08-31	/
/	/ PRIOR APPLICATION NUMBER: 60/058,612	/
/	/ PRIOR FILING DATE: 1997-09-10	/
/	/ NUMBER OF SEQ ID NOS: 58	/
/	/ SEQ ID NO: 13	/
/	/ SOFTWARE: Fast-SEQ for Windows Version 3.0	/
/	/ LENGTH: 29	/
/	/ TYPE: DNA	/
/	/ ORGANISM: H. sapiens	/
/	/ US-09-144-367-13	/

RESULT 2	US-0-9-372-339-2	TITLE OF INVENTION: CYP3A4 NFSE Variant and Methods of Use Therefor
/ Sequence 2, Application US/09372339	/ Patent No. 617484	/ GENERAL INFORMATION:
/	/ FILE REFERENCE: PENN-0059	/
/	/ CURRENT APPLICATION NUMBER: US/09/372,339	/
/	/ CURRENT FILING DATE: 1999-08-11	/
/	/ EARLIER APPLICATION NUMBER: 60/096,586	/
/	/ NUMBER OF SEQ ID NOS: 6	/
/	/ SOFTWARE: PatentIn Ver. 2.0	/
/	/ LENGTH: 1345	/
/	/ TYPE: DNA	/
/	/ ORGANISM: Homo sapiens	/
/	/ US-09-372-339-2	/

RESULT 5  
US-09-144-367-3

Query Match 91.1%; Score 16.4; DB 3; Length 1345;  
Best Local Similarity 94.4%; Pred. No. 34;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGACAGAG 18  
Db 806 GACAGGGCAGACAGAG 823

RESULT 3  
US-09-144-367-3

; Sequence 3, Application US/09144367  
; Patent No. 6432639  
; GENERAL INFORMATION:  
; APPLICANT: Licher, Jay  
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
; FILE REFERENCE: SEQ-12P  
; CURRENT APPLICATION NUMBER: US/09/144-367  
; CURRENT FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/058,612  
; PRIOR FILING DATE: 1997-09-10  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1345  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: Other  
; LOCATION: (0)...(0)  
; US-09-144-367-3.

Query Match 88.9%; Score 16; DB 4; Length 1345;  
Best Local Similarity 88.9%; Pred. No. 53;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGACAGAG 18  
Db 806 GACAGGGCAGACAGAG 823

RESULT 4  
US-09-144-367-44

; Sequence 44, Application US/09144367  
; Patent No. 6432639  
; GENERAL INFORMATION:  
; APPLICANT: Licher, Jay  
; APPLICANT: Guido, Marco  
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
; FILE REFERENCE: SEQ-12P  
; CURRENT APPLICATION NUMBER: US/09/144-367  
; CURRENT FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/058,612  
; PRIOR FILING DATE: 1997-09-10  
; NUMBER OF SEQ ID NOS: 58  
; SEQ ID NO 44  
; LENGTH: 19  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; US-09-144-367-44

Query Match 85.6%; Score 15.4; DB 4; Length 19;  
Best Local Similarity 94.1%; Pred. No. 70;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACAGGGCAGACAGAG 18  
Db 1 ACAGGGCAGACAGAG 17

RESULT 6  
US-09-144-367-12/C

; Sequence 6, Application US/09144367  
; Patent No. 6432639  
; GENERAL INFORMATION:  
; APPLICANT: Licher, Jay  
; APPLICANT: Guido, Marco  
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
; FILE REFERENCE: SEQ-12P  
; CURRENT APPLICATION NUMBER: US/09/144-367  
; CURRENT FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/058,612  
; PRIOR FILING DATE: 1997-09-10  
; NUMBER OF SEQ ID NOS: 58  
; SEQ ID NO 12  
; LENGTH: 32  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; US-09-144-367-12

Query Match 85.6%; Score 15.4; DB 4; Length 19;  
Best Local Similarity 94.1%; Pred. No. 70;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGACAGAG 18  
Db 27 GACAGGGCAGACAGAG 10

RESULT 7  
US-08-154-019-28/C

; Sequence 7, Application US/08154019  
; Patent No. 5633076  
; GENERAL INFORMATION:  
; APPLICANT: Deboer, Herman A.  
; APPLICANT: Stricker, Rein L.  
; APPLICANT: Heyneker, Herbert L.  
; APPLICANT: Lee, Sang He  
; APPLICANT: Pieper, Frank  
; APPLICANT: Krimpenfort, Paul J.A.  
;

```

TITLE OF INVENTION: Production of Recombinant Polypeptides
TITLE OF INVENTION: by Bovine Species and Transgenic Methods
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
    STREET: Townsend and Townsend and Crew
    CITY: One Market Plaza, Stewart Tower, Suite 2000
    STATE: California
    COUNTRY: USA
    ZIP: 94105

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/154,019
    FILING DATE: 16-NOV-1993
    CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/461,333
        FILING DATE: 05-JUN-1995
        APPLICATION NUMBER: US 08/077,788
        FILING DATE: 15-JUN-1993
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/895,956
        FILING DATE: 15-JUN-1992
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/619,131
        FILING DATE: 27-NOV-1990
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/444,745
        FILING DATE: 01-DEC-1989
ATTORNEY/AGENT INFORMATION:
    NAME: Liebeschuetz, Joe O.
    REGISTRATION NUMBER: 37,505
    REFERENCE/DOCKET NUMBER: 16994-003123
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 415-543-9600
    TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
    LENGTH: 807 base pairs
    TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-154-019-28

Query Match          82.2%; Score 14.8; DB 1; Length 807;
US-08-461-333-28/C Sequence 28, Application US/08461333
    General Information:
        Patient No.: 5741557
        Sequence: Deboer, Herman A.
        Applicant: Stricker, Rein
        Applicant: Reynaert, Herbert L.
        Applicant: Platenburg, Gerald
        Applicant: Lee, Sang He
        Applicant: Pieper, Frank
        Applicant: Krimpenfort, Paul J.A.
Title of Invention: Production of Recombinant Polypeptides
Title of Invention: by Bovine Species and Transgenic Methods
Number of Sequences: 38

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Stewart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105

; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,333
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,788
; FILING DATE: 15-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebesgrub, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-461-333-28

Query Match          82.2%; Score 14.8; DB 1; Length 9
Best Local Similarity 88.9%; Pred. No. 1.9e+02; Inde-
Matches 16; Conservate 0; Mismatches 2; Inde-
Qy      1 GACAGGGCAGGACAGAG 18
        ||||| | | | | | | |
Db      411 GACAGGGCAGGTCAAG 394

; RESULT 9
US-08-464-167-28/C
; Sequence 28, Application US/08464167
; Patient No. 601357
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Plattenburg, Gerald
; APPLICANT: Lee, Sang Hee
; APPLICANT: Bieper, Frank
; APPLICANT: Krimpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypep-
; TITLE OF INVENTION: by Bovine Species and Transgenic
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Stewart Tower, Suite 2000
; CITY: San Francisco
; STATE: California

```

COUNTRY: USA  
 ZIP: 94105  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/464,167  
 FILING DATE: 05-JUN-1995  
 CLASSIFICATION: 800  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/077,788  
 FILING DATE: 15-JUN-1993  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/895,956  
 FILING DATE: 15-JUN-1992  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/619,131  
 FILING DATE: 27-NOV-1990  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/444,745  
 FILING DATE: 01-DEC-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Liebeschuetz, Joe O.  
 REGISTRATION NUMBER: 37,505  
 REFERENCE/DOCKET NUMBER: 16994-003124  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-543-9600  
 TELEFAX: 415-543-5043  
 INFORMATION FOR SEQ ID NO: 28:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 807 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: Linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-464-167-28

Query Match 82.2%; Score 14.8; DB 3; Length 807;  
 Best Local Similarity 88.9%; Pred. No. 1.9e+02; Mismatches 0; Gaps 0;  
 Matches 16; Conservative 0; Indels 0;

QY 1 GACAAGGCGAGAGAG 18  
 ||||| ||||| ||||| |||||  
 Db 411 GACAGGAGACGGTCAAG 394

RESULT 11  
 US-08-464-167-28/C

Sequence 28, Application US/08476798  
 Patent No. 6066725

GENERAL INFORMATION:  
 APPLICANT: Deboer, Herman A.  
 Sequence 28, Application US/09158313  
 Patent No. 6066725

APPLICANT: Deboer, Herman A.  
 APPLICANT: Strijker, Rein  
 APPLICANT: Heyneker, Herbert L.  
 APPLICANT: Plattenburg, Gerald  
 APPLICANT: Lee, Sang He  
 APPLICANT: Pieper, Frank  
 APPLICANT: Krimpenfort, Paul J.A.

TITLE OF INVENTION: Production of Recombinant Polypeptides  
 TITLE OF INVENTION: by Bovine Species and Transgenic Methods  
 NUMBER OF SEQUENCES: 38  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew  
 STREET: One Market Plaza, Stewart Tower, Suite 2000  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94105

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/476,798  
 FILING DATE: 07-JUN-1995



APPLICATION NUMBER: US 07/619,131  
 FILING DATE: 27-NOV-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/444,745  
 FILING DATE: 01-DEC-1999  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Liebeschuetz, Joe O.  
 REGISTRATION NUMBER: 7,505  
 REFERENCE/DOCKET NUMBER: 16994-003123  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-543-9600  
 TELEFAX: 415-543-5043  
 INFORMATION FOR SEQ ID NO: 29:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 824 base pairs  
 TYPE: nucleic acid  
 SPANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 LENGTH: 824 base pairs  
 TYPE: nucleic acid  
 SPANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)

Query Match Score 14.8; DB 1; Length 824;  
 Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAGGGCAGGAGCAGAG 18  
 ||||| | | | | | | | | | | |  
 Db 417 GACAGGGCAGGAGCAGAG 400

RESULT 14  
 US-08-464-167-29/C  
 Sequence 29, Application US/08464167  
 Patent No. 6013857  
 GENERAL INFORMATION:  
 APPLICANT: Deboer, Herman A.  
 APPLICANT: Strijker, Rein  
 APPLICANT: Heyneker, Herbert L.  
 APPLICANT: Pieper, Frank  
 APPLICANT: Krimpenfort, Paul J.A.  
 TITLE OF INVENTION: Production of Recombinant Polypeptides  
 NUMBER OF SEQUENCES: 38  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew  
 STREET: One Market Plaza, Stewart Tower, Suite 2000  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94105  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/464,167  
 FILING DATE: 05-JUN-1995  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/077,788  
 FILING DATE: 15-JUN-1993  
 APPLICATION NUMBER: US 07/895,956  
 FILING DATE: 15-JUN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/619,131  
 FILING DATE: 27-NOV-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/444,745  
 FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:  
 NAME: Liebeschuetz, Joe O.  
 REGISTRATION NUMBER: 37,505  
 REFERENCE/DOCKET NUMBER: 16994-003124  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-543-9600  
 TELEFAX: 415-543-5043  
 INFORMATION FOR SEQ ID NO: 29:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 824 base pairs  
 TYPE: nucleic acid  
 SPANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-464-167-29  
 Query Match Score 82.2%; DB 3; Length 824;  
 Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GACANGGCAGGAGCAGAG 18  
 ||||| | | | | | | | | | | |  
 Db 417 GACAGGGCAGGAGCAGAG 400  
 RESULT 15  
 US-09-158-313-29/C  
 Sequence 29, Application US/09158313  
 Patent No. 6066725  
 GENERAL INFORMATION:  
 APPLICANT: Deboer, Herman A.  
 APPLICANT: Strijker, Rein  
 APPLICANT: Heyneker, Herbert L.  
 APPLICANT: Pieper, Frank  
 APPLICANT: Krimpenfort, Paul J.A.  
 TITLE OF INVENTION: Production of Recombinant Polypeptides  
 NUMBER OF SEQUENCES: 38  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew  
 STREET: One Market Plaza, Stewart Tower, Suite 2000  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94105  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/158,313  
 FILING DATE:  
 APPLICATION NUMBER: US 07/619,131  
 FILING DATE: 27-NOV-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/895,956  
 FILING DATE: 15-JUN-1992  
 APPLICATION NUMBER: US 07/444,745  
 FILING DATE: 01-DEC-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Liebeschuetz, Joe O.  
 REGISTRATION NUMBER: 37,505  
 REFERENCE/DOCKET NUMBER: 16994-003125  
 TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-543-9600  
; TELEFAX: 415-543-2043  
; INFORMATION FOR SEQ ID NO: 29:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 824 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-158-313-29

Query Match 82.2%; Score 14.8; DB 3; Length 824;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAGGCAGGACAG 18  
Db ||||| 1111 1111 1111  
Dy 417 GACAGGCAGGACAG 400

Search completed: August 27, 2003, 19:42:01  
Job time : 42.0385 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 17:59:19 : Search time 140:192 Seconds  
(without alignments) 346.595 Million cell updates/sec

Title: US-10-085-612-1.  
Perfect score: 18  
Sequence: 1 gacaaggccggcagacag 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512  
Minimum Match length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 100%  
Maximum Match 100%

Listing first 45 summaries  
1: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1980.DAT: \*  
2: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1982.DAT: \*  
3: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1982.DAT: \*  
4: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1983.DAT: \*  
5: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1984.DAT: \*  
6: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1985.DAT: \*  
7: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1986.DAT: \*  
8: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1987.DAT: \*  
9: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1988.DAT: \*  
10: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1989.DAT: \*  
11: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1990.DAT: \*  
12: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1991.DAT: \*  
13: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1992.DAT: \*  
14: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1993.DAT: \*  
15: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1994.DAT: \*  
16: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1995.DAT: \*  
17: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1996.DAT: \*  
18: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1997.DAT: \*  
19: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1998.DAT: \*  
20: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA1999.DAT: \*  
21: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA2000.DAT: \*  
22: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA2001A.DAT: \*  
23: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA2001B.DAT: \*  
24: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA2002.DAT: \*  
25: /SIDS1/gcdata/geneseq/geneseq/geneseq-emb1/NA2003.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	18	100.0	18	24	AAD45760	Human cytochrome P
2	16.4	91.1	22	24	AAD45777	Human promoter -39
c	3	16.4	91.1	29	AAX28306	PCR primer for Hum
c	4	16.4	91.1	1345	AAZ57020	Nucleic acid seque
	5	16.4	91.1	75235	ABV75557	Saccharopolyspora
	6	16.4	91.1	123785	ABX7717	DNA sequence of hu
	7	16.4	91.1	172637	AEN83124	Human voltage-acti
	8	16.4	91.1	237961	ABQ80552	Human CanIon gene

9	16	88.9	1345	20	AAX28296
10	16	88.9	3881	21	ABE68745
c	11	15.4	85.6	19	AAX28287
c	12	15.4	85.6	19	ABE68748
c	13	15.4	85.6	677	24
c	14	15.4	85.6	45546	AAZ23520
c	15	15.4	85.6	199161	ABE83564
c	16	15.4	83.3	779	ABE86612
c	17	15	83.3	779	ABZ74371
c	18	15	83.3	779	ABE67942
c	19	15	83.3	788	ABZ73544
c	20	15	83.3	788	ABE67148
c	21	15	83.3	788	ABE67148
c	22	15	83.3	224	ABE78354
c	23	15	83.3	3135	ABQ78354
c	24	15	83.3	3276	ABA58729
c	25	15	83.3	22	ABA199463
c	26	14.8	82.2	18	AAD45762
c	27	14.8	82.2	21	AAD36217
c	28	14.8	82.2	26	AAD45776
c	29	14.8	82.2	32	AAB28305
c	30	14.8	82.2	32	ABA50761
c	31	14.8	82.2	137	ABA14930
c	32	14.8	82.2	137	ABA109234
c	33	14.8	82.2	137	ABA42487
c	34	14.8	82.2	137	ABA165909
c	35	14.8	82.2	137	ABA126225
c	36	14.8	82.2	137	ABA17616
c	37	14.8	82.2	137	ABA#9225 used t
c	38	14.8	82.2	137	Human liver single
c	39	14.8	82.2	137	Human ovarian deriv
c	40	14.8	82.2	137	Human ORFX polyNUC
c	41	14.8	82.2	349	Novel human polynu
c	42	14.8	82.2	374	Novel human polynu
c	43	14.8	82.2	381	Human breast cell
c	44	14.8	82.2	429	Human foetal liver
c	45	14.8	82.2	429	Human foetal liver

### ALIGNMENTS

RESULT 1	ID AAD45760 standard	DNA: 18 BP.
	AC AAD45760;	
	XX	DT 27-DEC-2002 (first entry)
	XX	DE Human cytochrome P450 (CYP) 3A4 polymorphic variant DNA fragment.
	XX	XX Human; cytochrome P450; CYP3A4; CYP3A5; glutathione S-transferase; GSTM1; breast cancer; therapy; chemotherapeutic agent; variant; drug-drug interaction; drug adverse effect; anti-cancer agent; SNP; enzyme; single nucleotide polymorphism; ds.
	XX	OS Homo sapiens.
	XX	Synthetic.
	FH Key variation	Location/Qualifiers
	FT FT replace (11, A)	/standard_name= "Single nucleotide polymorphism (SNP)"
	FT FT /*tag= a	

PA (DNAS-) DNA SCI INC.  
 PA (UDU-) UNIV DUKE.  
 XX Guida M, Hall J, Petros WP, Vredenburgh JJ, Colvin OM, Marks JR;  
 PI DR WPI; 2002-691652/74.

XX New nucleic acid molecule useful for identifying polymorphisms associated with CYP3A4, CYP3A5 or GSTM1 substrates which influence breast cancer treatments, comprises at least one base variation from human CYP3A4 or CYP3A5 sequence -

PS Example 2; Page 26: 41pp; English.

XX The invention relates to a nucleic acid molecule comprising at least one base variation from human cytochrome P450 (CYP) 3A4 or CYP3A5 sequence. Nucleic acid molecules of the invention are useful for identifying polymorphisms associated with CYP3A4, CYP3A5 or glutathione S-transferase (GSTM1) substrates which influence breast cancer treatments. They are also useful in diagnostic purposes to identify individuals having a polymorphic genotype which influence the outcome of breast cancer treatments and the selection of chemotherapeutic agents used to treat breast cancer. The polymorphisms detected are used to screen altered metabolism of CYP3A4, CYP3A5 or GSTM1 substrates, potential drug-drug interactions, drug adverse effects, likelihood of successful clinical outcome following treatment with anti-cancer agents such as cisplatin, cyclophosphamide and/or BCNU. The present sequence is human promoter -392 CYP4503A4 specific probe.

XX Sequence 18 BP; 7 A; 3 C; 8 G; 0 U; 0 other;

XX Query Match Score 100.0%; Score 18; DB 24; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 32; Indels 0; Gaps 0;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 GACAAGGGAGGAGAG 18  
 ||||| | | | | | | | | | | |  
 Db 1 GACAAGGGAGGAGAG 18

RESULT 3  
 ID AAX28306/C  
 DE PCR primer for Human CYP3A4 gene promoter.  
 XX ID AAX28306 standard; DNA: 29 BP.  
 XX AC AAX28306;  
 XX DT 17-JUN-1999 (first entry)  
 XX DE PCR primer for Human CYP3A4 gene promoter.  
 XX ID AAX28306 standard; DNA: 29 BP.  
 XX KW CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism; CYP3A4 substrate; drug-drug interaction identification; toxin exposure; genetic linkage detection; phenotypic variation; promoter; PCR Primer;  
 XX KW ss.  
 XX DT 17-JUN-1999 (first entry)  
 XX OS Homo sapiens.  
 XX PN W09913105-A1.  
 XX DE Human promoter -392 CYP4503A4 specific probe #2.  
 XX DT 27-DEC-2002 (first entry)  
 XX KW Human; cytochrome P450; CYP3A4; CYP3A5; glutathione S-transferase;  
 KW polymorphism; GSTM1; breast cancer; therapy; chemotherapeutic agent;  
 KW drug-drug interaction; drug adverse effect; anti-cancer agent;  
 KW enzyme; probe; promoter; ss.  
 OS Homo sapiens.  
 XX PN WO200208448-A1.  
 XX PD 06-SEP-2002.  
 XX PF 26-FEB-2002; 2002WO-US06135.  
 XX PR 26-FEB-2001; 2001US-271630P.  
 XX PA (DNAS-) DNA SCI INC.  
 PA (UDU-) UNIV DUKE.  
 XX Guida M, Hall J, Petros WP, Vredenburgh JJ, Colvin OM, Marks JR;  
 PI DR WPI; 2002-691652/74.

XX New nucleic acid molecule useful for identifying polymorphisms

CC drug-drug interactions, and adverse/side effects as well as diseases that  
 CC result from environmental or occupational exposure to toxins. They can  
 CC also be used to establish animal, cell culture and *in vitro* cell-free  
 CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used  
 CC for expression studies to determine the effect of promoter and/or intron  
 CC sequence variations on mRNA expression and stability. These polymorphisms  
 CC are also used as single nucleotide polymorphisms to detect genetic  
 CC linkage to phenotypic variation in activity and expression of CYP3A4. The  
 CC nucleic acids can also be used to generate genetically modified non-human  
 CC animals or site specific gene modifications in cell lines.

XX Sequence 29 BP; 3 A; 12 C; 3 G; 11 T; 0 other;

Query Match 91.1%; Score 16.4; DB 20; Length 29;

Best Local Similarity 94.4%; Pred. No. 1; 8e-02;

Matches 17; Conservative 0; Mismatches 1;

Indels 0; Gaps 0;

Qy 1 GACAAGGGCAGGAGAG 18

Db ||||||| ||||| ||||| 823

RESULT 5 ABY75557 standard; DNA; 75236 BP.

XX ABY75557;

AC XX 22-JAN-2003 (first entry)

XX DT XX Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 1.

XX DE XX Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;

XX KW XX metabolite; spinosyn; gene; ds.

OS XX Saccharopolyspora sp.

OS XX

FH Key 1..13035

FT CDS /\*tag= a

FT /product= "busA"

FT FT 13059..19508

FT CDS /\*tag= b

FT /product= "busB"

FT FT 19553..29036

FT CDS /\*tag= c

FT /product= "busC"

FT FT 29092..43893

FT CDS /\*tag= d

FT /product= "busD"

FT FT 43945..60639

FT CDS /\*tag= e

FT /product= "busE"

FT FT 6200..63940

FT CDS /\*tag= f

FT /product= "ORF RI"

FT FT /transl\_except= (pos:62879..62881,aa:Xaa)

FT FT /note= "No start codon given"

FT CDS /note= "XaaUnknown, No start codon given"

FT FT complement (65226..66602)

FT FT /\*tag= g

FT /product= "ORF RII"

FT FT /note= "No start codon given"

FT CDS complement (68759..69676)

FT FT /\*tag= h

FT /product= "ORF RIII"

FT FT /note= "No start codon given"

XX WO200009752-A1.

XX PD 10-OCT-2002.

XX XX

XX PR 28-MAR-2002; 2002WO-US0968.

XX PR 30-MAR-2001; 2001US-280175P.

XX PA (DOW ) DOW AGROSCIENCES LLC.

XX PI Hahn DR, Jackson JD, Bullard BS, Gustafson GP, Waldron C;

XX PI Mitchell JC;

XX DR WPI; 2003-058434/05;

DR P-BSD; ABP57679, ABP57680, ABP57681, ABP57682, ABP57705,

DR ABP57706, ABP57707.

XX PT New butenyl-spinosyn biosynthetic genes, useful for increasing the  
 PT production of butenyl-spinosyn insecticidal macrolides, or for changing

SQ Sequence 29 BP; 3 A; 12 C; 3 G; 11 T; 0 other;

Query Match 91.1%; Score 16.4; DB 21; Length 1345;

Best Local Similarity 94.4%; Pred. No. 2.2e+02;



```

FT variation replace (8145,A)
FT FT /*tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (8999,T)
FT FT /*tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (9044,C)
FT FT /*tag= o
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (9990,A)
FT FT /*tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (1012,T)
FT FT /*tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (11855,C)
FT FT /*tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (12177,G)
FT FT /*tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (12348,T)
FT FT /*tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (13446,A)
FT FT /*tag= u
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (16751,C)
FT FT /*tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (20631,G)
FT FT /*tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (20587,G)
FT FT /*tag= x
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (20733,A)
FT FT /*tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (20935,G)
FT FT /*tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT variation 23450..23533
FT FT /*tag= aa
FT /number= 3
FT intron 23534..25047
FT FT /*tag= ab
FT /number= 3
FT exon 25048..25187
FT FT /*tag= ac
FT /number= 4
FT intron 25188..25275
FT FT /*tag= ad
FT /number= 4
FT variation 25276..25404
FT FT /*tag= ae
FT /number= 5
FT intron 25405..55496
FT FT /*tag= af
FT /number= 5
FT variation replace (27330,C)
FT FT /*tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (27602,T)
FT FT /*tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (27615..27616,G)
FT FT /*tag= aj
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace (30002,G)
FT FT /*tag= bi

```

```

FT variation /standard-name="Single nucleotide polymorphism"
FT replace ("75509,T")
FT /*tag= bj

Query Match 91.1%; Score 16 4; DB 24; Length 172637;
Best local Similarity 94.4%; Pred. NO. 2.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gap 0

QY 1 GACAGGGGAGGAGAG 18
    ||||||| | | | | |
Db 95503 GACAAGGGAGGGAGAG 955520

RESULT 8
ABQ080552 standard; DNA; 237961 BP.
XX ID ABQ080552
XX AC ABQ080552;
XX DT 08-NOV-2002 (first entry)
XX DE Human Canion gene fragment #2.
XX KW Human; Canion; neuroleptic; hypotensive; anticonvulsant; analgesic;
XX KW antiarrhythmic; antiangular; cardiotonic; antidiuressant;
XX KW gene therapy; schizophrenia; voltage-gated ion channel; bipolar disorder;
XX KW central nervous system disorder; cardiovascular disorder; hypertension;
XX KW pain; epilepsy; chromosome 13q; gene; ds.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX exon 43726..43866
XX FT /*tag= a
XX FT /number= 8
XX exon 43998..44102
XX FT /*tag= b
XX FT /number= 9
XX misc_difference 51090
XX FT /*tag= c
XX FT /note= "Biallelic marker: 99-79335-60/A7"
XX exon 52093..52179
XX FT /*tag= d
XX FT /number= 10
XX misc_difference 61293
XX FT /*tag= e
XX FT /note= "Biallelic marker: 99-79336-369/A8"
XX exon 77568..77699
XX FT /*tag= f
XX FT /number= 11
XX misc_difference 80602
XX FT /*tag= g
XX FT /note= "Biallelic marker: 99-79338-332/A9"
XX exon 98226..98393
XX FT /*tag= h
XX FT /number= 12
XX misc_difference 100485
XX FT /*tag= i
XX FT /note= "Biallelic marker: 99-79314-201/A10"
XX misc_difference 100509
XX FT /*tag= j
XX FT /note= "Biallelic marker: 99-79314-225/A11"
XX exon 106567..106758
XX FT /*tag= k
XX FT /number= 13
XX misc_difference 106725
XX FT /*tag= l
XX FT /note= "Biallelic marker: 99-79316-158/A12"
XX exon 14109..144246
XX FT /*tag= m
XX FT /number= 14
XX exon 15919..159668
XX FT /*tag= n

```

CC that block Cation channels are useful to treat a number of diseases or  
 CC conditions, preferably schizophrenia or bipolar disorder, and also  
 CC including pain disorders, epilepsy and various cardiovascular disorders  
 CC such as heart arrhythmias, angina and hypertension. The Cation gene is  
 CC located on chromosome 13q. The present sequence comprises a genomic  
 CC sequence of Cation, comprising the exons 8 to 27.

XX Sequence 237961 BP; 74252 A; 46176 G; 72420 T; 1307 other;

RESULT 9  
QY AAX28296 standard; DNA; 1345 BP.

ID ABK68745 standard; DNA; 1345 BP.

DB ABK68745 standard; DNA; 1345 BP.

XX XX

AC AC

XX XX

PD PD

XX XX

07 -MAR-2002.

DT 17-JUN-1999 (first entry)

XX XX

DE DE

XX XX

DE Human CYP3A4 gene promoter.

XX XX

KW CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;

KW CYP3A4 substrate; drug-drug interaction identification; toxin exposure;

KW genetic linkage detection; phenotypic variation; promoter; ss.

XX XX

PI PI

XX XX

Risinger C, Andersson MK, Lewander T, Olaission E;

XX XX

DR DR

XX XX

WPI; 2002-351712/38.

PN PN

PR PR

XX XX

PT PT

XX XX

PT PT

XX XX

PA PA

XX XX

(GEMI-) GEMINI GENOMICS PLC.

XX XX

PI PI

XX XX

Risinger C, Andersson MK, Lewander T, Olaission E;

XX XX

DR DR

XX XX

WPI; 2002-351712/38.

PN PN

PR PR

XX XX

PT PT

XX XX

PT PT

XX XX

PA PA

XX XX

(AXYS-) AXYS PHARM INC.

XX XX

PI PI

XX XX

Guida M, Lichter JB;

XX XX

DR DR

WPI; 1999-215070/18.

XX XX

PT PT

XX XX

New isolated CYP3A4 polymorphic sequences

XX XX

PS PS

02-SEP-1998;

98WO-US18158.

XX XX

PR PR

10-SEP-1997;

97US-0058812.

XX XX

PA PA

XX XX

(AXYS-) AXYS PHARM INC.

XX XX

PI PI

XX XX

Guida M, Lichter JB;

XX XX

DR DR

WPI; 1999-215070/18.

XX XX

PT PT

XX XX

New isolated CYP3A4 polymorphic sequences

XX XX

PS PS

Disclosure; Page 29; 40pp; English.

XX XX

This sequence represents the human CYP3A4 gene promoter.

CC The invention relates to a CYP3A4 sequence polymorphism,

CC which is part of a non-naturally occurring chromosome. Nucleic acids

CC comprising the CYP3A4 polymorphic sequences can be used to screen

CC patients for altered metabolism for CYP3A4 substrates, potential

CC drug-drug interactions, and adverse/side effects as well as diseases that

CC can also be used to establish animal, cell culture and in vitro cell-free

CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used

CC for expression studies to determine the effect of promoter and/or intron

CC sequence variations on mRNA expression and stability. The polymorphisms

CC are also used as single nucleotide polymorphisms to detect genetic

CC linkage to phenotypic variation in activity and expression of CYP3A4. The

CC nucleic acids can also be used to generate genetically modified non-human

CC animals or site specific gene modifications in cell lines.

XX Sequence 1345 BP; 352 A; 314 C; 356 G; 322 T; 1 other;

Query Match 88.9%; Score 16; DB 20; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;

Best Local Similarity 88.9%; Pred. No. 3.4e+02;

Mismatches 1; Indels 0; Gaps 0;

XX Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;  
Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DE Human secreted protein gene 10 clone HDGP94.  
 XX  
 KW Human; secreted protein; fusion protein; gene; therapy;  
 KW protein therapy; diagnosis; tissue; cancer; tumour; AIDS;  
 KW autoimmune disorder; allergy; cardiovascular; viral; bacterial;  
 KW fungal infection; immunosuppressive; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20029422-A1.  
 XX  
 PD 25-MAY-2000.  
 XX  
 PT 09-NOV-1999; 99WO-US26409.  
 XX  
 PR 12-NOV-1998; 98US-0108207.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ni J., Ruben SM, Rosen CA, Ebner R, Florence KA, Young PE;  
 BIrc GE, Carter KC, Komatsu S G,  
 DR WPI; 2000-387729/33.  
 XX  
 PT Novel human secreted proteins useful for diagnosing, preventing,  
 PT ameliorating a medical condition e.g. cardiovascular  
 PT disease -  
 XX  
 PS Claim 1; Page 233-234; 295pp; English.  
 XX  
 CC The present sequence represents a nucleic acid molecule which encodes a  
 CC secreted human protein. The gene number and the clone it was derived  
 CC from are given in the descriptor line.  
 CC The invention relates to 31 novel genes and their fragments (nucleic  
 CC acid sequences: AAA61260-A61293; amino acid sequences AB12201-B12371)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new  
 CC polynucleotides. Specific uses are described for each of the 31  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC and include products for the diagnosis or treatment of cancer, tumours,  
 CC AIDS, autoimmune disorders, allergy, cardiovascular disorders, viral,  
 CC bacterial and fungal infection. The genes are used to generate fusion  
 CC proteins by linking to the gene a human immunoglobulin portion (AAA1251)  
 CC for increasing stability of the fused protein as compared to the  
 CC secreted protein only.  
 XX  
 Sequence 3881 BP; 1043 A; 716 C; 1512 T; 0 other;  
 XX  
 Query Match Score 88.9%; Score 16; DB 21; Length 3881;  
 Best Local Similarity 100 %; Pred. No. 3.6e+02; Indels 0; Gaps 0;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 Qy 3 CAACGGCAGGACAGAG 18  
 AC ||||||| ||||| ||||| |||||  
 XX  
 ID AAX28287 standard; DNA; 19 BP.  
 XX  
 AC AAX28287;  
 XX  
 DT 17-JUN-1999 (first entry)  
 XX  
 DE Human CYP3A4 gene polymorphism #1.  
 XX  
 KW CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;  
 KW CYP3A4 substrate; drug-drug interaction; identification; toxin exposure;  
 KW genetic linkage detection; phenotypic variation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9913106-A1.  
 XX  
 PD 18-MAR-1999.  
 XX  
 PT 02-SEP-1998; 98WO-US18158.  
 XX  
 PR 10-SEP-1997; 97US-0058612.  
 XX  
 PA (AXYS) AXYS PHARM INC.  
 XX  
 PI Guida M, Lichter JB;  
 XX  
 DR WPI; 1999-215070/18.  
 XX  
 PT New isolated CYP3A4 polymorphic sequences  
 XX  
 PS Claim 2; Page 35; 40pp; English.  
 XX  
 CC This sequence represents a CYP3A4 sequence polymorphism of the invention,  
 CC which is part of a non-naturally occurring chromosome. Nucleic acids  
 CC comprising the CYP3A4 polymorphic sequences can be used to screen  
 CC patients for altered metabolism for CYP3A4 substrates, potential  
 CC drug interactions, and adverse/side effects as well as diseases that  
 CC result from environmental or occupational exposure to toxins. They can  
 CC also be used to establish animal, cell culture and in vitro cell-free  
 CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used  
 CC for expression studies to determine the effect of promoter and/or intron  
 CC sequence variations on mRNA expression and stability. The polymorphisms  
 CC are also used as single nucleotide polymorphisms to detect genetic  
 CC linkage to phenotypic variation in activity and expression of CYP3A4. The  
 CC nucleic acids can also be used to generate genetically modified non-human  
 CC animals or site specific gene modifications in cell lines.  
 XX  
 SQ Sequence 19 BP; 7 A; 3 C; 9 G; 0 U; 0 other;  
 XX  
 Query Match Score 85.6%; Score 15.4%; DB 20; Length 19;  
 Best Local Similarity 94.1%; Pred. No. 5.3e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 Qy 2 ACAAGGGAGGAGAG 18  
 AC ||||||| ||||| ||||| |||||  
 XX  
 Db 1 ACAAGGGAGGAGAG 17  
 XX  
 RESULT 13  
 ABK68748 standard; DNA; 19 BP.  
 XX  
 ID ABK68748;  
 AC ABK68748;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide #2 for detecting polymorphism in CYP3A4 gene.  
 XX  
 KW Human; single nucleotide polymorphism; SNP; cytochrome p450; CYP;  
 KW CYP3A4; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218641-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PT 30-AUG-2001; 2001WO-IB01580.  
 XX  
 PR 30-AUG-2000; 2000GB-0021286.  
 XX  
 PA (GEMI-) GEMINI GENOMICS PLC.  
 XX  
 RI Risinger C, Andersson MK, Lewander T, Olaisson E;

DR WPI; 2002-351712/38.  
 XX Novel primer pairs and sequence determination oligonucleotides useful  
 PT for amplifying and detecting novel single nucleotide polymorphisms in  
 PT the 5'-flanking regions of cytochrome p450 (CYP)3A4 and CYP2C9 genes  
 PT respectively -  
 XX Disclosure; Page 3; 47pp; English.  
 XX The present invention relates to PCR primer pairs for amplifying  
 CC and sequence determination oligonucleotides for detecting single  
 CC nucleotide polymorphisms (SNPs) in the 5'-flanking regions of human  
 CC cytochrome p450 (CYP) genes encoding CYP3A4 or CYP2C9. The SNPs  
 CC correspond to position 461 of a 1345 base pair sequence for  
 CC CYP3A4 or position 957, 1049, 1164, 1526, 1661 and 1662 of a 2438 base  
 CC pair sequence for CYP2C9. The PCR primers are useful for amplifying  
 CC the CYP sequences and the oligonucleotides are useful for detecting  
 CC SNPs in the 5'-flanking regions of the CYP3A4 or CYP2C9 genes.  
 CC ABK68747-ABK68750 represent previous published oligonucleotides  
 CC for detecting a polymorphism in the CYP3A4 gene.

XX Sequence 19 BP; 7 A; 3 C; 9 G; 0 U; 0 other;

Query Match 85.6%; Score 15.4; DB 24; Length 19;  
 Best Local Similarity 94.1%; Pred. No. 5.3e+02;  
 Matches 16; Conservative 0; Indels 0; Gaps 0;  
 PR 2 ACATGGGAGGACAGAG 18  
 DB 1 ACATGGGAGGAGGAGG 17

RESULT 14

ABQ71128 ID ABQ71128 standard; DNA; 677 BP.  
 XX ABQ71128;  
 AC XX DT 29-AUG-2002 (first entry)  
 DE Listeria monocytogenes 4b specific contig55.  
 XX KW Antibacterial; Listeria; food contamination; mutational analysis;  
 KW infection; ds.  
 OS XX PD 11-APR-2002.  
 XX Listeria monocytogenes 4b.  
 XX PN WO200228891-A2.  
 XX PD 04-OCT-2001; 2001WO-PRO3061.  
 XX PR 04-OCT-2000; 2000FR-0012697.  
 XX PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Kunst F; Glaser P;  
 DR WPI; 2002-332479/37.

New genomic sequences from Listeria species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators -  
 XX Claim 22; SEQ ID 3941; 180pp; French.  
 PS The present invention relates to nucleic acid sequences  
 CC (AB067188 ABQ7112) from Listeria sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of Listeria (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be

CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of Listeria (potential therapeutic agents), also for  
 CC treating infections by Listeria, and are useful as immunogens in  
 CC anti-Listeria vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at [ftp.wipo.int/pub/published\\_pot\\_sequences](http://ftp.wipo.int/pub/published_pot_sequences).

XX SQ Sequence 677 BP; 145 A; 133 C; 206 G; 193 T; 0 other;

Query Match 85.6%; Score 15.4; DB 24; Length 677;

Best Local Similarity 94.1%; Pred. No. 6.2e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACAAGGGCACGACAGAG 18

DB 60 AGAAGGGCAGGACAGAG 76

RESULT 15

AAX23520 ID AAX23520 standard; DNA; 45546 BP.  
 XX AC AAX23520;  
 XX DT 23-JUN-1999 (first entry)

DE Human kidney aminopeptidase P genomic DNA fragment 4.  
 XX KW Aminopeptidase; human; Amp; gene therapy; treatment; Amp-deficiency;  
 KW prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;  
 KW arterial stenosis; industrial protein feed; malabsorption syndrome;  
 KW proteinaceous waste degradation; additive; immunohistochemistry; ss;  
 KW Homo sapiens.  
 XX OS Homo sapiens.  
 XX PN WO9911799-A2.  
 XX PD 11-MAR-1999.  
 XX PR 02-SEP-1998; 98WO-US18426.  
 XX PR 02-SEP-1997; 97US-0057854.  
 XX PA (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.

XX PI Ryan JW; Sprinkle RJJC; Venema RC;  
 XX DR WPI; 1999-205193/17.

Nucleic acid encoding human aminopeptidase P  
 PT  
 XX PS Claim 13; Page 165-192; 201pp; English.

XX This invention describes the isolation of a novel human aminopeptidase P  
 CC (Amp). This protein is used to produce recombinant Amp and can be used  
 CC for gene therapy for treating Amp-deficiency conditions. Its fragments and  
 CC are used as primers and probes to identify patients with homozygous and  
 CC heterozygous Amp deficiency, including prenatal diagnosis (patients  
 CC defective in Amp are at risk of developing angioidema if treated with  
 CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors  
 CC in cases of excessive Amp expression. The product of the invention is  
 CC also used to identify Amp-expressing sequences in other animals and to  
 CC generate transgenic animals, and comparisons of genomic sequences are  
 CC used to detect mutations. Amp inhibitors are potentially useful as  
 CC antihypertensive agents and to prevent or treat arterial (restenosis  
 CC or atherosclerosis). The structure of Amp is used to design synthetic  
 CC substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal  
 CC imido bonds, can be used to degrade industrial protein feeds to free  
 CC amino acids, to degrade proteinaceous wastes, as additives in enzyme  
 CC formulations used to treat malabsorption syndrome and for studying its  
 CC biological role. Antibodies against Amp are used in immunohistochemical  
 CC methods to study Amp distribution.

xx	SQ	Sequence 45546 BP; 12027 A; 11359 C; 11380 G; 10780 T; 0 other;
	Query Match	Score 15.4;
	Best Local Similarity	94.1%; Pred. No. 7 6e+02;
	Matches	16; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
QY	1	GACAGGGCAGGACAGA 17
Dy	565	GGCAAGGGCAGGACAGA 581

Search completed: August 27, 2003, 18:15:44  
Job time : 142.192 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

Nucleic acid nucleic search, using sw model

August 27, 2003, 18:02:39 ; Search time 69.9615 Seconds  
(without alignments)

214.504 Million cell updates/sec

total number of hits satisfying chosen parameters: 1139956

Score	No.	Query Match	Length	DB ID	Description
1	1	22.8	67.1	1345 3	US-09-372-339-1
2	2	22.8	67.1	1345 3	US-09-372-339-2
3	3	22.8	67.1	1345 4	US-09-372-339-3
4	4	20	58.8	357 2	US-08-394-143-1
5	5	20	58.8	357 3	US-09-393-831-1
6	6	20	58.8	357 4	US-09-393-831-1
c	7	20	58.8	829 2	US-08-394-143-3
c	8	20	58.8	829 3	US-08-394-143-3
c	9	20	58.8	829 4	US-09-393-831-3
c	10	19.2	56.5	34185 4	US-09-394-143-3
c	11	19	55.9	4550 3	US-09-393-831-3
c	12	19	55.9	4550 4	US-09-393-831-3
c	13	19	55.9	4960 4	US-09-393-831-3
c	14	18.8	55.3	238 1	US-07-903-46-32
c	15	18.8	55.3	238 2	PCT-US93-05794-32
c	16	18.8	55.3	3117 3	US-08-364-268-20
c	17	18.8	55.3	3117 4	US-09-234-146-20
c	18	18.8	55.3	1767 1	US-07-903-46-2
c	19	18.8	55.3	1767 5	PCT-US93-05794-2
c	20	18.8	55.3	2487 4	US-09-394-143-19
c	21	18.8	55.3	2772 4	US-09-257-894-12
c	22	18.8	55.3	3018 1	US-07-903-46-1
c	23	18.8	55.3	3018 5	PCT-US93-05794-1
c	24	18.8	55.3	15664 1	US-08-402-222-3
c	25	18.8	55.3	15664 1	US-08-508-004-3
c	26	18.8	55.3	15664 1	US-08-404-066-3
c	27	18.8	55.3	15664 1	US-08-402-008-3

RESULT 1  
US-09-372-339-1

; Sequence 1, Application US/09372339

; Patent No. 6174684

; GENERAL INFORMATION:

; APPLICANT: Rebbeck, Timothy

; INVENTOR: Felix, Carolyn

; TITLE OF INVENTION: CYP3A4 NFSE Variant and Methods of Use Therefor

; FILE REFERENCE: PENN-09695

; CURRENT APPLICATION NUMBER: US/09-372-339

; CURRENT FILING DATE: 1999-08-11

; EARLIER APPLICATION NUMBER: 60/096,586

; EARLIER FILING DATE: 1998-08-14

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1345

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-372-339-1

RESULT 2  
US-09-372-339-2

; Sequence 1, Application US/09372339

; Patent No. 6174684

; GENERAL INFORMATION:

; APPLICANT: Rebbeck, Timothy

; INVENTOR: Felix, Carolyn

; TITLE OF INVENTION: CYP3A4 NFSE Variant and Methods of Use Therefor

; FILE REFERENCE: PENN-095

; CURRENT APPLICATION NUMBER: US/09-372-339

; CURRENT FILING DATE: 1999-08-11

; EARLIER APPLICATION NUMBER: 60/096,586

; EARLIER FILING DATE: 1998-08-14

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 1345

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-372-339-2

RESULT 3  
US-09-372-339-3

; Sequence 1, Application US/09372339

; Patent No. 6174684

; GENERAL INFORMATION:

; APPLICANT: Rebbeck, Timothy

; INVENTOR: Felix, Carolyn

; TITLE OF INVENTION: CYP3A4 NFSE Variant and Methods of Use Therefor

; FILE REFERENCE: PENN-095

; CURRENT APPLICATION NUMBER: US/09-372-339

; CURRENT FILING DATE: 1999-08-11

; EARLIER APPLICATION NUMBER: 60/096,586

; EARLIER FILING DATE: 1998-08-14

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 1345

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-372-339-3

SUMMARIES

Score	No.	Query Match	Length	DB ID	Description
1	1	22.8	67.1	1345 3	Sequence 1, Appli
2	2	22.8	67.1	1345 3	Sequence 2, Appli
3	3	22.8	67.1	1345 4	Sequence 3, Appli
c	4	20	58.8	357 2	Sequence 1, Appli
c	5	20	58.8	357 3	Sequence 1, Appli
c	6	20	58.8	357 4	Sequence 1, Appli
c	7	20	58.8	829 2	Sequence 3, Appli
c	8	20	58.8	829 3	Sequence 3, Appli
c	9	20	58.8	829 4	Sequence 3, Appli
c	10	19.2	56.5	34185 4	Sequence 3, Appli
c	11	19	55.9	4550 3	Sequence 35, Appli
c	12	19	55.9	4550 4	Sequence 3, Appli
c	13	19	55.9	4960 4	Sequence 3, Appli
c	14	18.8	55.3	238 1	Sequence 32, Appli
c	15	18.8	55.3	238 2	Sequence 32, Appli
c	16	18.8	55.3	3117 3	Sequence 20, Appli
c	17	18.8	55.3	3117 4	Sequence 20, Appli
c	18	18.8	55.3	1767 1	Sequence 2, Appli
c	19	18.8	55.3	1767 5	Sequence 2, Appli
c	20	18.8	55.3	2487 4	Sequence 19, Appli
c	21	18.8	55.3	2772 4	Sequence 12, Appli
c	22	18.8	55.3	3018 1	Sequence 1, Appli
c	23	18.8	55.3	3018 5	Sequence 1, Appli
c	24	18.8	55.3	15664 1	Sequence 3, Appli
c	25	18.8	55.3	15664 1	Sequence 3, Appli

RESULTS

Query Match 67.1%; Score 22.8; DB 3; Length 1345;  
 Best Local Similarity 92.3%; Pred. No. 4.2;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 9 TGCTACTGGCTGAGGTGAGCCCG 34  
 Db 1035 TGCTACTGGCTGAGGTGAGCCCG 1060

## RESULT 3

US-09-144-367-3

Sequence 3, Application US/09144367

Patent No. 643239

GENERAL INFORMATION:

APPLICANT: Lichten, Jay

TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4

FILE REFERENCE: SEQ-12P

CURRENT APPLICATION NUMBER: US/09/144,367

CURRENT FILING DATE: 1998-08-31

PRIOR APPLICATION NUMBER: 60/058,612

PRIOR FILING DATE: 1997-09-10

NUMBER OF SEQ ID NOS: 58

SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 3

LENGTH: 1345

TYPE: DNA

ORGANISM: H. sapiens

PRATURE:

NAME/KEY: Other

LOCATION: (0) . . . (0)

Query Match 67.1%; Score 22.8; DB 4; Length 1345;  
 Best Local Similarity 92.3%; Pred. No. 4.2;  
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 9 TGCTACTGGCTGAGGTGAGCCCG 34  
 Db 1035 TGCTACTGGCTGAGGTGAGCCCG 1060

## RESULT 4

US-08-294-143-1/c

Sequence 1, Application US/08294143

Patent No. 5874221

GENERAL INFORMATION:

APPLICANT: NAHUM SONENBERG

APPLICANT: ARNIM PAUSE

APPLICANT: JOE B. HARFORD

APPLICANT: VINCENT J. MILES

TITLE OF INVENTION: METHODS FOR TREATING

TITLE OF INVENTION: HORMONE DISORDERS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon &amp; Lyon

STREET: 633 West Fifth Street

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

MEDIUM TYPE: storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294,143

FILING DATE: August 22, 1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA: including application  
 PRIORITY APPLICATION DATA: described below:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warburg, Richard J.  
 REGISTRATION NUMBER: 3,2,327  
 REFERENCE/DOCKET NUMBER: 202/060  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEFAX: (213) 955-0440  
 TELEX: 67-3510  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 357 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-294-143-1

Query Match 58.8%; Score 26; DB 2; Length 357;  
 Best Local Similarity 82.1%; Pred. No. 41;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 CTTCGCTACTGGCTGCAAGCTGCAGCCC 33  
 Db 35 CTTCGGCTGGCTGAGTGCTGCC 8

RESULT 5

US-09-256-331-1/c

Sequence 1, Application US/09256331

PATENT INFORMATION:  
 Patent No. 6411077

GENERAL INFORMATION:

APPLICANT: NAHUM SONENBERG  
 APPLICANT: ARNIM PAUSE  
 APPLICANT: JOE B. HARFORD  
 APPLICANT: VINCENT J. MILES  
 TITLE OF INVENTION: METHODS FOR TREATING  
 TITLE OF INVENTION: HORMONE DISORDERS  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Lyon & Lyon  
 STREET: 633 West Fifth Street  
 STREET: Suite 4700  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90071

COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 MEDIUM TYPE: storage  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/256,331  
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/294,143  
 FILING DATE: August 22, 1994

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEFAX: (213) 955-0440  
 TELEX: 67-3510  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 357 base pairs  
 TYPE: nucleic acid

;

;

STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-593-483-1

Query Match 58.8%; Score 20; DB 3; Length 357;  
Best Local Similarity 82.1%; Pred. No. 41;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTTGCTACTGGCAGCTGAGCC 33  
Db 35 CTGGGTCTGGTCACTGCCTGCC 8

RESULT 6  
US-09-593-483-1/c  
Sequence 1, Application US/09593483  
Patent No. 641075  
GENERAL INFORMATION:  
APPLICANT: NAHUM SONENBERG  
ARTIM PAUSE  
TITLE OF INVENTION: METHODS FOR TREATING  
HORMONE DISORDERS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,143  
FILING DATE: August 22, 1994  
CLASSIFICATION: A24  
PRIORITY APPLICATION DATA:  
PRIORITY APPLICATION NUMBER: US/08/294,143  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 202/060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 829 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-294-143-3

;

;

Query Match 58.8%; Score 20; DB 2; Length 829;  
Best Local Similarity 82.1%; Pred. No. 48;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTTGCTACTGGCTGGAGCTGCAGCCCC 33  
Db 62 CTGGGTCTGGTCACTGCCTGCC 35

RESULT 8  
US-09-256-331-3/c  
Sequence 3, Application US/09256331  
Patent No. 611077  
GENERAL INFORMATION:  
APPLICANT: ARMIN PAUSE  
APPLICANT: JOE B. HARFORD  
APPLICANT: VINCENT J. MILES  
TITLE OF INVENTION: METHODS FOR TREATING  
HORMONE DISORDERS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700

;

;

Query Match 58.8%; Score 20; DB 4; Length 357;  
Best Local Similarity 82.1%; Pred. No. 41;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTTGCTACTGGCTGGAGCTGCAGCCCC 33  
Db 35 CTGGGTCTGGTCACTGCCTGCC 8

RESULT 7

CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90071  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 COMPUTER: IBM Compatible  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/256,331  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/294,143  
 FILING DATE: August 22, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warburg, Richard J.  
 REGISTRATION NUMBER: 32,327  
 REFERENCE/DOCKET NUMBER: 202/060  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEX: (213) 955-0440  
 TELEFAX: 67-3510  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 829 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09-593-483-3

Query Match 58.8%; Score 20; DB 3; Length 829;  
 Best Local Similarity 82.1%; Pred. No. 48;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 CTTTGCTACTGGCTGAGCTGCAGGCC 33  
 Db 62 CTTGGGTCTGGCTGAGCTGTGCC 35

RESULT 9  
 US-09-593-483-3/c  
 Sequence 3, Application US/09/593483  
 Patent No. 6110715  
 GENERAL INFORMATION:  
 APPLICANT: NAHUM SONENBERG  
 ARNIN PAUSE  
 JOE B. HARFORD  
 VINCENT J. MILES  
 TITLE OF INVENTION: METHODS FOR TREATING  
 HORMONE DISORDERS  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Lyon & Lyon  
 STREET: 633 West Fifth Street  
 SUITE 4700  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90071  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 storage  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/593,483  
 FILING DATE: 14-Jun-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/294,143  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warburg, Richard J.  
 REGISTRATION NUMBER: 32,327  
 REFERENCE/DOCKET NUMBER: 202/060  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEX: 67-3510  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 829 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09-593-483-3  
 US-09-545-481-3  
 Sequence 3, Application US/09/545481  
 Patient No. 6451319  
 GENERAL INFORMATION:  
 APPLICANT: Chiang, Christina H.  
 APPLICANT: Cochran, Mark D.  
 PATENT NO. 6451319  
 FILE REFERENCE: SY093K US  
 CURRENT APPLICATION NUMBER: US/09/545,481  
 CURRENT FILING DATE: 2000-04-07  
 PRIOR APPLICATION NUMBER: 60/128,766  
 PRIOR FILING DATE: 1999-04-09  
 NUMBER OF SEQ ID NOS: 17  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 3  
 LENGTH: 34185  
 TYPE: DNA  
 ORGANISM: Bovine adenovirus type 1  
 US-09-545-481-3

Query Match 56.5%; Score 19.2; DB 4;  
 Best Local Similarity 75.0%; Pred. No. 1.8e-02;  
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 GATTCTTGTACTGGCTGAGCTGCAGGCC 33  
 Db 12572 GATTATTCACAGGGCCGCTGTAGCC 12603

RESULT 11  
 US-09-313-294A-2795  
 Sequence 2795, Application US/09/313294A  
 Patient No. 6476212  
 GENERAL INFORMATION:  
 APPLICANT: Laljudi, Raghunath V.  
 APPLICANT: Ito, Laura Y.  
 PATENT NO. 6476212  
 FILE REFERENCE: BL-0017 US  
 CURRENT APPLICATION NUMBER: US/09/313,294A  
 CURRENT FILING DATE: 1999-05-14  
 NUMBER OF SEQ ID NOS: 7600  
 SOFTWARE: PERL Program  
 SEQ ID NO 2795

```

; LENGTH: 286
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700553275H1
US-09-313-294A-2795

Query Match 55.9%; Score 19; DB 4; Length 286;
Best Local Similarity 81.5%; Pred. No. 97; Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 3 ATCTTGCTTACGGCTGAGTCAGTGAG 29
Db 17 ATCTTGCGAGCTGAGCTGAG 43

RESULT 12
US-09-103-663-35/C
Sequence 35, Application US/09103663D
; Patent No. 617103
; GENERAL INFORMATION:
; APPLICANT: Kinet et al.
; TITLE OF INVENTION: Isolation, characterization, and use of the human beta
; TITLE OF INVENTION: subunit of the high affinity receptor for
; TITLE OF INVENTION: immunoglobulin E. SOFTWARE: Patentin Ver. 2.1
; FILE REFERENCE: 50490
; CURRENT APPLICATION NUMBER: US/09/103,663D
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 07/869,933
; EARLIER FILING DATE: 1992-04-16
; NUMBER OF SEQ ID NOS: 35
; SEQ ID NO: 35
; LENGTH: 4550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; LOCATION: (578)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; LOCATION: (735)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; LOCATION: (1362)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; LOCATION: (2479)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; LOCATION: (2517)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; LOCATION: (2526)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; LOCATION: (2549)..(2552)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; LOCATION: (2612)..(2614)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; NAME/KEY: misc_feature
; LOCATION: (2633)
; OTHER INFORMATION: n represents a, c, t or g.

; FEATURE: misc_feature
; LOCATION: (2920)
; OTHER INFORMATION: n represents a, c, t or g.
; FEATURE: misc_feature
; LOCATION: (3069)
; OTHER INFORMATION: n represents a, c, t or g.

; FEATURE: misc_feature
; LOCATION: (4348)
; OTHER INFORMATION: (572)...(4348)

RESULT 13
US-09-907-843-3
Sequence 3, Application US/0907843
; Patent No. 6440339
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE 2 EXPRESSION
; FILE REFERENCE: RITS-0279
; CURRENT APPLICATION NUMBER: US/09/907,843
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 87
; SEQ ID NO: 3
; LENGTH: 4960
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (572)...(4348)
US-09-907-843-3

Query Match 55.9%; Score 19; DB 4; Length 4960;
Best Local Similarity 81.5%; Pred. No. 1.5e+02; Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 4 TTCCTTGCTACTGGCTCAGTGAGC 30
Db 383 TCCCTTGCTGGCGCGAGCG 409

RESULT 14
US-07-903-466-32/C
Sequence 32, Application US/07903466
; Patent No. 5395767
; GENERAL INFORMATION:
; APPLICANT: Painter, Robert B.
; APPLICANT: Painter, Robert B.
; APPLICANT: Murane, John P.
; APPLICANT: Yu, Loh C.
; TITLE OF INVENTION: Gene for Ataxia-Telangiectasia
; TITLE OF INVENTION: Complementation Group D (ATDC)
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: Stewart Street Tower, 18th Fl., One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: San Francisco
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 FILING DATE: 19920622  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: 91-077-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-777-9275  
 TELEFAX: 415-543-4219  
 INFORMATION FOR SEQ ID NO: 32:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 238 base pairs  
 TYPE: NUCLEAR ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-07-303-166-32

Query Match 55.3%; Score 18.8; DB 1; Length 238;  
 Best Local Similarity 76.7%; Pred. No. 1.1e+02;  
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CGATCTTGTACTGGTGAGCTGAGCG 30  
 |||| | | | | | | | | | | | | | | | | | | | |  
 Db 145 CGACCGTTGCTCTGAGGCATCTGAGC 116

RESULT 15  
 PCT-US93-05794-32/C  
 Sequence 32, Application PC/US93/05794  
 GENERAL INFORMATION:  
 APPLICANT: The Regents of the University of California  
 TITLE OF INVENTION: Gene for Ataxia-Telangiectasia  
 TITLE OF INVENTION: Complementation Group D (ATDC)  
 NUMBER OF SEQUENCES: 45  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Leona L. Lauder  
 STREET: 177 Post Street, Suite 800  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: San Francisco  
 ZIP: 94108-4731  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/05794  
 FILING DATE: 19930618  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/903,466  
 FILING DATE: 22-JUN-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lauder, Leona L.  
 REGISTRATION NUMBER: 30,863  
 REFERENCE/DOCKET NUMBER: 91-077-1 PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-421-4973  
 TELEFAX: 415-421-1663  
 INFORMATION FOR SEQ ID NO: 32:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 238 base pairs  
 TYPE: nuclear acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO

ALIGNMENTS

University of Washington  
Box 35145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066657244  
Email: raymond@u.washington.edu  
Class: shotgun  
Location: classifiers  
FEATURES:





/lab\_host="SOLR"  
 /clone.lib="GM Drosophila melanogaster ovary BlueScript"  
 /note="Organ: ovary; Vector: BlueScript SK; Site:1: ECR1;  
 Site:2: XhoI; Constructed using Stratagene ZAP-CDNA  
 Synthesis kit. Oligo dT-primed and directionally cloned at  
 BASE COUNT 139 a 154 c 172 g 138 t

Query Match 65.3%; Score 22.2; DB 9; Length 603;  
 Best Local Similarity 88.9%; Pred. No. 8.9e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATTCTTGCTACTGGCTGCACTGG 27  
 Db 262 CGATTCGGCTGGCTACTGGCTGCACTGG 288

RESULT 7 AA538677 LOCUS AA538677\_2 GI:13767084 EST.  
 DEFINITION LD18261\_5prime LD Drosophila melanogaster embryo Bluescript  
 Drosophila melanogaster cDNA clone LD18261\_5 similar to CalpB;  
 FBan0008107 calpain, located on: 3L 67C3-67C3; 04/10/2001, mRNA  
 sequence. AA538677  
 ACCESSION AA538677  
 VERSION 1  
 KEYWORDS EST.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Phylloidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 61).  
 Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
 Lewis,S and Rubin,G.M.  
 TITLE BDGP/HMI Drosophila EST Project  
 JOURNAL Unpublished  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2794820.  
 Other\_Esns: LD18261\_3prime  
 Contact: Stapleton, M.  
 BDGP Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu  
 hit genomic AB00350: arm.3L [698134, 9979619]  
 estimated cyto:67C1-67C9: 04/10/2001  
 Plate: LD182 row: F column: 1  
 High quality sequence stop: 500  
 POLYA-No.

FEATURES Source  
 1. 611  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="BDGP EST:BDG1n017400"  
 /db\_xref="BDGP EST:BDG1n017400"  
 /clone="LD18261"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: embryo; Vector: Bluescript SK; Site:1: ECR1;  
 Site:2: XhoI; Constructed using Stratagene ZAP-CDNA  
 Synthesis kit. Oligo dT-primed and directionally cloned at  
 ECR1 and XhoI in BlueScript SK(+/-)"

BASE COUNT 134 a 197 c 158 g 122 t

ORIGIN

Query Match 65.3%; Score 22.2; DB 9; Length 611;  
 Best Local Similarity 88.9%; Pred. No. 9e+02;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

---

QY 1 CGATCTTGCTACTGGCTGCACTGG 27  
 Db 545 CGATTCGGCTACTGGCTGCACTGG 571

RESULT 8 BH613513 LOCUS BH613513  
 DEFINITION SALK\_034371 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
 thaliana genomic clone SALK\_034371, genomic survey sequence.  
 BH613513  
 GSS: Arabidopsis thaliana (thale cress)

ACCESSION BH613513..1 GI:18062637

KEYWORD SOURCE ; eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 ORGANISM ; eurosid II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab  
 ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prendis,L., Shinn,P.,  
 Zimmerman,J. and Ecker,J.R.  
 TITLE A Sequence-Indexed Library of Insertion Mutations in the  
 Arabidopsis Genome  
 Unpublished  
 Contact: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6319  
 Email: ecker@salk.edu  
 This is single pass sequence recovered from the left border of  
 TDNA. This sequence lies within an annotated exon of At1g20480.  
 Class: TDNA tagged.

JOURNAL  
 COMMENT  
 FEATURES source  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="Taxon:3702"  
 /clone.lib="SALK\_034371"  
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines" each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tDNA\\_protocols.html](http://signal.salk.edu/tDNA_protocols.html)"

LOCATION/QUALIFIERS  
 1. 180  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="Taxon:3702"  
 /clone.lib="SALK\_034371"  
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines" each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tDNA\\_protocols.html](http://signal.salk.edu/tDNA_protocols.html)"

BASE COUNT 42 a 45 c 38 g 55 t

ORIGIN

Query Match 64.7%; Score 22; DB 28; Length 180;  
 Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 TCTTGCCTACTGGCTGCACTGGCTGCAGCCCG 145  
 Db 116 TTTCGCTATTGTGAAGCTGAGCCGG 145

RESULT 9 BH214265 LOCUS BH214265  
 DEFINITION SALK\_010491 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
 thaliana genomic clone SALK\_010491, genomic survey sequence.  
 BH214265  
 VERSIN BH214265..1 GI:16396850  
 GSS:  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

REFERENCE 1 (bases 1 to 196) eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Geddrinab  
 C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prendis,L., Shinn,P.  
 Zimmerman,J., and Ecker,J.R.  
 TITLE A Sequence-Indexed Library of Insertion Mutations in the  
 Arabidopsis Genome  
 JOURNAL Unpublished  
 COMMENT Contact: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 558 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@salk.edu  
 This is single pass sequence recovered from the left border of  
 TDNA. This sequence lies within an intron of At1g20480.  
 Class: TDNA tagged.

FEATURES source  
 Location/Qualifiers  
 1..196  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="#taxon:3702"  
 /clone\_lib="SARK\_010491"  
 /clone\_id="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines  
 each of which contains one or more TDNA insertion  
 elements. The resultant fragment for each line was  
 directly sequenced to determine the genomic sequence at  
 the site of insertion. Details of the protocols used can  
 be found at [http://signal.salk.edu/tDNA\\_protocols.html](http://signal.salk.edu/tDNA_protocols.html)"

BASE COUNT ORIGIN  
 43 a 48 c 47 g 58 t

Query Match Best Local Similarity 64.7%; Score 22; DB 28; Length 196;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RESULT 11  
 At478360  
 LOCUS At478360 480 bp mRNA linear ;EST 24-MAY-2002  
 DEFINITION At478360 S00011 Hordeum vulgare cDNA clone S000110014G04F1, mRNA  
 sequence.  
 ACCESSION At478360  
 VERSION At478360.1 GI:21194515  
 KEYWORDS EST,  
 SOURCE Hordeum vulgare  
 ORGANISM Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 ; Triticeae; Hordeum.  
 1 (bases 1 to 480)  
 REFERENCE Saren,A.-M.; Tanskanen,J., Paulin,L. and Schulman,A.H.  
 AUTHORS Barley EST's  
 TITLE Unpublished  
 JOURNAL Institute of Biotechnology  
 COMMENT University of Helsinki  
 P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014  
 Finland.

FEATURES source  
 Location/Qualifiers  
 1..480  
 /organism="Hordeum vulgare"  
 /mol\_type="mRNA"  
 /db\_xref="#taxon:513"  
 /clone="S000110014G04F1"  
 /dev\_stage="Developing seed"  
 /clone\_id="S00011"  
 /note="12,15-18 days after pollination"

BASE COUNT ORIGIN  
 94 a 171 c 137 g 78 t

Query Match Best Local Similarity 64.1%; Score 21.8; DB 9;  
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;



Db 536 |||||||GCTTCTCCCTCTGGTGGAGTGCACTCG 504

RESULT 15

CNS04FD7/c CNS04FD7 904 bp DNA linear GSS 01-SEP-2000

DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 106E22 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL288196

VERSION AL288196.1 GI:8026722

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraontoidea; Tetraontidae; Tetraodon.

REFERENCE 1 Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Windre, P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

AUTHORS Estimate of human gene number provided by genome-wide analysis

TITLE using Tetraodon nigroviridis DNA sequence

JOURNAL Nat. Genet. 25 (2), 235-238 (2000)

MEDLINE 20296613

PUBMED 1083565

REFERENCE 2 Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

AUTHORS Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

TITLE Genome Res. 10 (7), 939-945 (2000)

JOURNAL MEDLINE 2059837

PUBMED 10899143

REFERENCE 3 (bases 1 to 904)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr

- Web : [www.genoscope.cns.fr/](http://www.genoscope.cns.fr/))

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES Source

ORGANISM "Tetraodon nigroviridis"

/mol\_type="Genomic DNA"

/db\_xref="taxon:99883"

/clone\_id="G"

/note="Genoscope sequence ID : COBG106FC11LP1-end : T7"

BASE COUNT 237 a 210 c 236 g 217 t 4 others

ORIGIN

Query Match 64.1%; Score 21.8; DB 29; Length 904;

Best Local Similarity 78.8%; Pred. No. 1.4e+03;

Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CGATTCTTTGCTACTGGCTGAGTCAGCTGCC 33

Db 648 CGTTCTGCCAACGGCTGAGCTGAGCCCC 616

Search completed August 27, 2003, 19:40:02  
Job time : 2276.12 secs

Gencore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:00:44 ; Search time 536.538 Seconds  
(without alignments)  
1372.452 Million cell updates/sec

Title: US-10-085-612-1

Perfect score: 18

Sequence: 1 gacaaggccggacagacag 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing First 45 summaries

Database : GenEmbl:\*

1: gb\_ba;\*  
2: gb\_hhg;\*  
3: gb\_in;\*  
4: gb\_cm;\*  
5: gb\_ov;\*  
6: gb\_pat;\*  
7: gb\_ph;\*  
8: gb\_pl;\*  
9: gb\_pr;\*  
10: gb\_ro;\*  
11: gb\_sus;\*  
12: gb\_sy;\*  
13: gb\_an;\*  
14: gb\_y1;\*  
15: em\_be;\*  
16: em\_fun;\*  
17: em\_hum;\*  
18: em\_in;\*  
19: em\_mu;\*  
20: em\_cm;\*  
21: em\_or;\*  
22: em\_ov;\*  
23: em\_pat;\*  
24: em\_ph;\*  
25: em\_p1;\*  
26: em\_o;\*  
27: em\_sts;\*  
28: em\_un;\*  
29: em\_v1;\*  
30: em\_htg\_hum;\*  
31: em\_htg\_irv;\*  
32: em\_htg\_other;\*  
33: em\_htg\_mus;\*  
34: em\_htg\_p1n;\*  
35: em\_htg\_r0d;\*  
36: em\_htg\_men;\*  
37: em\_htg\_vrt;\*  
38: em\_sy;\*  
39: em\_htg\_o\_hum;\*  
40: em\_htg\_o\_mus;\*  
41: em\_htg\_o\_other;\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

	Result No.	Score	Query %	Match	Length	DB	ID	Description
	1	18	100.0	144869	2	AC114884		AC114884 Rattus no
	2	18	100.0	178857	2	AC114881		AC114881 Rattus no
	3	18	100.0	188490	2	AC118733		AC118733 Mus muscu
	4	18	100.0	196662	2	AC12737		AC12737 Rattus no
	5	18	100.0	209706	2	AC101931		AC101931 Mus muscu
	6	18	100.0	217691	2	AC103052		AC103052 Rattus no
	7	18	100.0	246620	2	AC119882		AC119882 Mus muscu
	8	17	94.4	47589	2	AC124053		AC124053 Mus muscu
	9	17	94.4	116608	2	AC121356		AC121356 Mus muscu
	10	17	94.4	151613	2	AC144886		AC144886 Bos tauru
	11	17	94.4	163475	10	AC026761		AC026761
	12	17	94.4	173362	2	AC141188		AC141188 Bos tauru
	13	17	94.4	196351	2	AC134942		AC134942 Mus muscu
	14	17	94.4	205621	10	AC079082		AC079082 Mus muscu
	15	17	94.4	217181	2	AC132033		AC132033 Rattus no
	16	17	94.4	219937	2	AC135670		AC135670 Mus muscu
	17	17	94.4	237330	2	AC098104		AC098104 Rattus no
	18	17	94.4	238658	10	AC079818		AC079818 Mus muscu
	19	17	94.4	239553	2	AC123201		AC123201 Rattus no
	20	17	94.4	271890	2	AC106243		AC106243 Rattus no
	21	17	94.4	287536	2	AC093993		AC093993 Rattus no
	22	17	94.4	347924	2	AC131229		AC131229 Mus muscu
	23	16.4	91.1	29	6	AR222903		AR222903 Sequence
	24	16.4	91.1	532	4	AJ29058		AJ29058 Capra hircus
	25	16.4	91.1	573	9	CH2292058		CH2292058
	26	16.4	91.1	573	9	HS295982		HS295982 Homo sapi
	27	16.4	91.1	590	11	G95069		G95069 S20962B/RF
	28	16.4	91.1	1345	6	AR142140		AR142140 Sequence
	29	16.4	91.1	8088	4	CBHLG		CBHLG
	30	16.4	91.1	32735	9	AC006124		AC006124 C. hircus ge
	31	16.4	91.1	38016	9	AC138433		AC138433 Homo sapi
	32	16.4	91.1	38235	9	AC004559		AC004559 Homo sapi
	33	16.4	91.1	60271	2	AC110286		AC110286 Homo sapi
	34	16.4	91.1	62915	2	AI929091_16		AI929091_16 Continuation (17 o
	35	16.4	91.1	64705	2	AC136358		AC136358 Homo sapi
	36	16.4	91.1	75236	6	AX600586		AX600586 Sequence
	37	16.4	91.1	75761	9	AL359733		AL359733 Human DNA
	38	16.4	91.1	88592	10	AL732630		AL732630 Mouse DNA
	39	16.4	91.1	103706	9	AL554891		AL554891 Human DNA
	40	16.4	91.1	104871	9	AC005484		AC005484 Homo sapi
	41	16.4	91.1	106041	10	AC074559		AC074559 Mus muscu
	42	16.4	91.1	108195	9	AC074129		AC074129 Homo sapi
	43	16.4	91.1	110000	2	AC106311		AC106311 Rattus no
	44	16.4	91.1	118172	9	AC015912		AC015912 Continuation (13 o
	45	16.4	91.1	123778	9	AC069294		AC069294 Homo sapi

## ALIGNMENTS

## RESULT 1

AC114884  
LOCUS

Rattus norvegicus clone RP31-557J15 strain Brown Norway, WORKING  
DRAFT SEQUENCE, 8 ordered pieces.

AC114884  
VERSION

HTG ; FPGS\_PHASE ; HTGS\_DRAFT.

Rattus norvegicus (Norway rat)

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Rattus.  
1 (bases 1 to 144869)

REFERENCE AUTHORS  
Akhter,N., Antonellis,A., Ayelle,K., Beckstrom-Sternberg,S.M.,

**JOURNAL** Benjamin,B., Blakesley,R.W., Bouffard,G., Brinkley,C., Brooks,S., Cariglia,K., Coleman,B., Dietrich,N.L., Granite,S., Guan,X., J.R., Gupta,J., Haghjhi,P., Han,J.J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,O.L., Maduro,V.B., Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,L.J., Pauqirigan,C., Pearson,R., Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C., Scantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L., Walker,M., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

**REFERENCE** NISC Comparative Sequencing Initiative

**AUTHORS** Unpublished

**TITLE** Direct Submission

**JOURNAL** Submitted [13-MAR-2002] NIH Intramural Sequencing Center, 8717

**COMMENT** Grovemont Circle, Gaithersburg, MD 20877, USA

**AUTHORS** Green,E.D.

**TITLE** Direct Submission

**JOURNAL** Submitted [25-SEP-2002] NIH Intramural Sequencing Center, 8717

**COMMENT** Grovemont Circle, Gaithersburg, MD 20877, USA  
On Sep 25, 2002 this sequence version replaced gi:19387611.

**AUTHORS**

**TITLE** NIH Intramural Sequencing Center

**JOURNAL** Center code: NISC

**COMMENT** Web site: <http://www.nisc.nih.gov>  
Contact: nisc.zo@nigri.nih.gov  
Project Information  
Center Project name: crxn  
Center clone name: 557315

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

**SEQUENCING** Sequencing vector: Plasmid; n/a; 108 of reads

**ASSEMBLY** Assembly program: Phrap; version 0.990319

**CONSENSUS** Consensus quality: 1431282 bases at least Q40

**CONSENSUS** Consensus quality: 143743 bases at least Q30

**CONSENSUS** Consensus quality: 144038 bases at least Q20

**INSERT** Insert size: 151000; agarose-fp

**QUALITY** Insert size: 144169; sum-of-contigs

**QUALITY** Quality coverage: 10.51x in Q20 bases; agarose-fp

**QUALITY** Quality coverage: 11.01x in Q20 bases; sum-of-contigs

**STATISTICS** Summary Statistics

**NOTE** This is a 'working draft' sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submittor.

This sequence will be replaced

- \* by the finished sequence as soon as it is available and the accession number will be preserved.
- \* 1 48966: contig of 48966 bp in length
- \* 1 49066: gap of unknown length
- \* 49067 74022: contig of 24956 bp in length
- \* 74023 74122: gap of unknown length
- \* 74123 76637: contig of 2515 bp in length
- \* 76638 76731: gap of unknown length
- \* 76738 8224: contig of 5787 bp in length
- \* 82255 82264: gap of unknown length

**FEATURES**

**source** 1. 144869

**/organism="Rattus norvegicus"**

**/mol\_type="genomic DNA"**

**/strain="Brown Norway"**

**/db\_xref="Daxton:10116"**

**/clone="RP31-557J15"**

**/clone\_lib="RP31"**

**misc\_feature** 1. 59406

**/note="clone overlaps with GenBank Accession Number AC114881 clone RP31-385022 (center project name cxo)"**

**misc\_feature** 1. 144866

**/note="assembly\_fragment**

**clone\_end:17**

**vector\_side:left"**

**misc\_feature** 49067 . 74022

**/note="assembly\_fragment"**

**misc\_feature** 74123 . 76637

**/note="assembly\_fragment"**

**misc\_feature** 76738 . 8224

**/note="assembly\_fragment"**

**misc\_feature** 82265 . 99342

**/note="assembly\_fragment"**

**misc\_feature** 99443 . 110906

**/note="assembly\_fragment"**

**misc\_feature** 110907 . 1121200

**/note="assembly\_fragment"**

**misc\_feature** 121301 . 144869

**/note="assembly\_fragment"**

**misc\_feature** 121301 . 144869

**clone\_end:56**

**vector\_side:right"**

**BASE COUNT** 39556 a 35181 C 34455 g 34977 t 700 others

**ORIGIN**

**Query Match** 100.0%; Score 18; DB 2; Length 144869;

**Best Local Similarity** 100.0%; Pred. No. 67; Mismatches 0; Indels 0; Gaps 0;

**Matches** 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**ACCESSION** AC114881

**VERSION** AC114881.2 GI:21629279

**DEFINITION** Rattus norvegicus clone RP31-385022 strain Brown Norway, WORKING DRAFT SEQUENCE, 4 ordered pieces.

**KEYWORDS** HTGS PHASE2; HTGS\_DRAFT.

**ORGANISM** Rattus norvegicus

**MATERIALS** Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

**REFERENCE** Akhter,N , Antonellis,A , Ayel, K , Beckstrom-Sternberg,S.M , Benjamin,B , Blakesley,R.W , Bouffard,G.G , Brein,K , Brinkley,C , Brooks,S , Dietrich,N.L , Granite,S , Guan,X , Gupta,J , Haghjhi,P , Hansen,N , Ho,S.-L , Idol,J.R , Karlins,E , Laric,P , Lee-Jin,S -Q , Legaspi,R , Maduro,Q.I , Maduro,V.B , Margulies,E.H , Masiello,C , Mastriani,S.D , McCloskey,J.C , McDevell,J , Pearson,R , Portnoy,M.E , Prasad,A , Schueler,M.G , Stantripop,S , Thomas,J.W , Thomas,P.J , Touchman,J.W , Tsugeon,C , Vogt,J.L , Walker,M.A ,

Wetherby,K.D., Wiggins,J., Young,A., Zhang,L.-H. and Green,E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 178857)

AUTHORS  
Green,E.D.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (13-MAR-2002) NIH Intramural Sequencing Center, 8717  
Groenveld Circle, Gaithersburg, MD 20877, USA  
COMMENT  
3 (bases 1 to 178857)  
AUTHORS  
Green,E.D.  
TITLE  
Direct Submission  
Groenveld Circle, Gaithersburg, MD 20877, USA  
On Jun 29, 2002 this sequence version replaced gi:19387608.  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: nisc.zoonthuri.nih.gov  
Project Information  
Center project name: oxo  
Center clone name: 385022

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Assembly: Dye-terminator Big Dye; 100% of reads  
Consensus Quality: 178011 bases at least Q40  
Consensus Quality: 178364 bases at least Q30  
Insert size: 178496 bases at least Q20  
Insert size: 178557; sum-of-contigs  
Quality coverage: 11.61x in Q20 bases; agarose-fp  
Quality coverage: 11.61x in Q20 bases; sum-of-contigs  
Insert size: 178557; sum-of-contigs  
Quality coverage: 11.61x in Q20 bases; agarose-fp  
Quality coverage: 11.61x in Q20 bases; sum-of-contigs  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces of the gaps between them are based on estimates that have provided by the submitter.  
\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 97908; contig of 97908 bp in length  
\* 97908 9806; gap of unknown length  
\* 97909 141249; contig of 43241 bp in length  
\* 141250 141349; gap of unknown length  
\* 141350 168491; contig of 27142 bp in length  
\* 168492 168591; gap of unknown length  
\* 168592 178557; contig of 10266 bp in length  
FEATURES  
Source  
Location/Qualifiers  
1. 178857  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="Brown Norway"  
/db\_xref="Taxon:10116"  
/clone\_id="RP1-385022"  
/note="assembly-fragment"

clone\_end\_SP6  
vector\_size:left"  
misc\_feature 98009 .111249  
/note="assembly\_fragment"  
2 119530 .178857  
misc\_feature  
/note="clone overlaps with GenBank Accession Number AC114884 clone RP31-57J15 (center project name cxn)"  
141350 .16491  
misc\_feature 168592 .178857  
/note="assembly\_fragment"  
clone\_end\_T7  
vector\_size:right"  
misc\_feature 47760 a 42335 c 42544 g 45938 t 300 others  
ORIGIN  
Query Match 100.0%; Score 18; DB 2; Length 178857;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
AC118733 188490 bp DNA Linear HTG 22-FEB-2003  
LOCUS clone RP24-170A20 WORKING DRAFT SEQUENCE, 3 unordered  
DEFINITION Mus musculus clone AC118733  
PIECES  
Mus musculus clone AC118733.3 GI:28467248  
HTG; HTGS PHASE; HTGS DRAFT.  
Eukaryota; Metazoa;  
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
ORGANISM  
Mus musculus (house mouse)  
Unpublished  
Birren,B., Linton,L., Gusbaum,C., and Lander,E.  
1 (bases 1 to 188490)  
REFERENCE  
AUTHORS Birren,B., Linton,L., Gusbaum,C., and Lander,E.  
TITLE  
JOURNAL  
2 (bases 1 to 188490)  
Unpublished  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Bouslikha,L.,  
Boukhaguet,P., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazarro,B., Chospel,Y., Collange,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,R., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goeyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Larocque,J., Lamazares,R.,  
Landers,J., Leroozky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marcus,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Meheus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicoll,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Doneill,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phankhang,P., Pierre,N., Polliard,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rose,C., Rosov,P.,  
Roman,J., Rosetti,M., Roy,J., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Strange,S., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talanay,J.J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Yaman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembok,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
3 (bases 1 to 188490)  
Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE  
AUTHORS Birren,B., Gusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavsky,L., Boughalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corrum,B., Dearallano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,

**TITLE** Direct Submission  
**JOURNAL** Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141 USA  
**COMMENT** On Feb 22, 2003 this sequence version replaced gi:28209705.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1999-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

**Center:** Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 -----, Project Information  
 Center project name: L25151  
 Center clone name: 170A-20

**Sequencing Vector:** Plasmid; n/a; 100% of reads  
**Chemistry:** Dye-terminator Big Dye; 100% of reads  
**Assembly program:** Phrap; version 0.960731  
**Consensus quality:** 187861 bases at least Q40  
**Consensus quality:** 188043 bases at least Q30  
**Consensus quality:** 188126 bases at least Q20  
**Insert size:** 183000; agarose-fp  
**Insert size:** 18290; sum-of-contigs  
**Quality coverage:** 11.4 in Q20 bases; agarose-fp  
**Quality coverage:** 11.1 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 \* 1 9810: contig of 9810 bp in length  
 \* 9811 9910: gap of 100 bp  
 \* 9911 113719: contig of 103809 bp in length  
 \* 113720 113819: gap of 100 bp  
 \* 113820 188490: contig of 74671 bp in length.

**FEATURES**  
**source** Location/Qualifiers  
 1..9810  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10190"  
 /clone="RP24-170A0"  
 /clone.lib="RPCI-24 Male Mouse BAC"  
 1..9810  
 /note="assembly\_fragment  
 clone\_end=sp6  
 vector\_side:left"  
 9911..113719  
 /note="assembly\_fragment  
 113820..188490  
 /note="assembly\_fragment  
 clone\_end=T"  
 vector\_side:right"  
 200 others

**misc\_feature**

**BASE COUNT** 57443 a 38263 c 37627 g 54957 t

**ORIGIN**  
**Query Match Score** 100.0%  
**Best Local Similarity** 100.0%  
**Matches** 18; **Conservative** 0;  
**Mismatches** 0; **Indels** 0; **Gaps** 0;

Qy 1 GACAGGGCAGGAGAG 18  
 Db 28760 GACAGGGCAGGAGAG 28743

**RESULT** 4  
**LOCUS** AC129737.3 GI:22856305  
**DEFINITION** Rattus norvegicus clone CH230-50651, \*\*\* SEQUENCING IN PROGRESS

**ACCESSION** AC129737  
**VERSION** 3  
**KEYWORDS** HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ENRICHED.

**SOURCE** Rattus norvegicus (Norway rat)

**ORGANISM** Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Rattus.

**REFERENCE** 1 (bases 1 to 196662)  
 Muzny,D., Marie,, Meckler,M., Lee,, Abramzon,S., Adams,C., Alder,J., Anil,A., Allen,H., Asbrooks,S., Amin,A., Anil,D., Arraldepechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Bitwali,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhal,B., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesario,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chiu,J., Cleveland,C., Cockrell,R., Cox,C., Coyne,M., Cree,A., D'Souza,L., Davis,M.L., Davis,C., Davy-Carroll,L., De Andia,C., Dederich,D., Delgado,O., Deason,S., Dermo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,R., Duval,B., Eaves,K., Egan,A., Escott,C., Eugene,C., Evans,C.A., Faiss,T., Fan,G., Fernandes,S., Finley,M., Flagg,N., Forbes,J., Foster,M., Foster,P., Fraser,C.M., Gabisa,A., Gaitta,R., Garcia,A., Garner,T., Gerza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guarratne,P., Haaland,W., Hamill,C., Hamilton,C., Hernandez,R., Hines,S., Hladun,S.L., Hodges,A., Hogan,A., Hollins,B., Howell,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,J., Jiang,H., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Keily,S., King,L., Kovar,C., Li,J., Lewis,L., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Loresuherwa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindarine,M., Mahmoud,M., Mallojy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawlinney,S., McLeod,M.P., McNeill,T.Z., Meinen,B., Milosavljevic,A., Miner,G., Milja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Norris,S., Nwakeleneh,O., Okuyonu,G., Olarinpaasoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Snee,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sora,J., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Puato,M., Quiror,J., Rachlin,E., Reeves,K., Reier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Snee,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sora,J., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wright,J., Warren,R., Wei,X., White,F., Williams,G., Wilson,R., Wleczik,R., Woeden,H., Worley,K., Wright,J., Wright,R., Wu,J., Yaku,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

Weinstock, G. and Gibbs, R.A.		Db	114096 GACAGGGAGGAGAG 114079
Direct Submission			
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 196662)		
AUTHORS	Worley, K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
COMMENT	3 (bases 1 to 196662)		
AUTHORS	Rat Genome Sequencing Consortium.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	On Sep 14, 2002 this sequence version replaced gi:22538766.		
AUTHORS	The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both ends sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.		
TITLE	Genome Center		
COMMENT	Center: Baylor College of Medicine		
AUTHORS	Center code: BCM		
REFERENCE	Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>		
AUTHORS	Contact: hgsc-help@bcm.tmc.edu		
TITLE	Project Information		
COMMENT	Center project name: GWEK		
AUTHORS	Center clone name: CH230-506G1		
TITLE	Summary Statistics		
COMMENT	Assembly program: Phrap; version 0.990329		
AUTHORS	Consensus quality: 18533 bases at least Q40		
REFERENCE	Consensus quality: 186194 bases at least Q30		
AUTHORS	Consensus quality: 180040 bases at least Q20		
COMMENT	Estimated insert size: 20426; sum-of-contigs estimation		
AUTHORS	Quality coverage: 4x in Q20 bases; sum-of-contigs estimation		
FEATURES	* NOTE: Estimated insert size may differ from sequence length. * (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html</a> ). * NOTE: This is a working draft sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submittor. * This sequence will be replaced. * by the finished sequence as soon as it is available and the accession number will be preserved. * 1. 196662: contig of 196662 bp in length.		
COMMENT	Location/Qualifiers		
AUTHORS	1. 1.196662 "Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-506G1"		
COMMENT	misc_feature 1..2028 /note="wgs-contig"		
BASE COUNT	51858 a 40862 c 42866 g 54119 t 6957 others		
ORIGIN			
COMMENT	Query Match 100.0%; Score 18; DB 2; Length 196662; Best Local Similarity 100.0%; Pred. No. 65; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 GACAGGGAGGAGAG 18 		

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center project name: L17758  
 Center clone name: 63\_0\_23

\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter.  
 \* This sequence will be replaced by the finished sequence as soon as it is available and by the accession number will be preserved.  
 \* The accession number will be 209706: contig of 209706 bp in length.

FEATURES  
 Source Qry 1 GRANAGGCCAGAGCAG 18  
 Db 32614 GACAGGGCAGGACAGAG 32631

BASE COUNT ORIGIN  
 AC103052/C 65611 a 41280 c 40670 g 62145 t

RESULT  
 AC103052 6  
 LOCUS 217691 bp DNA linear HTG 13-MAY-2003  
 DEFINITION Rattus norvegicus clone CH230-147P15, WORKING DRAFT SEQUENCE.  
 ACCESSION AC103052  
 VERSION AC103052.5 GI:30580047  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Mammalia; Eutheria; Rodentia; Chordata; Vertebrata; Euteleostomi;  
 Rattus.  
 REFERENCE 1 (bases 1 to 217691)  
 AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,Y., Aoyagi,A., Ayodeji,M., Baca,F., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswalo,K., Blair,J., Blankenburg,K., Calderon,E.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z.,  
 Cockrell,C., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.I., Davis,C., Davy-Carroll,L., De Andra,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escott,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garcia,M.,  
 Gebregiorgis,E., Geer,K., Gil,I.R., Grady,M., Guerra,W., Guerara,W.,  
 Harvey,Y., Haylak,P., Hayes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.I., Hodgson,A., Hogue,M.,  
 Hollins,B., Howell,S., Hulyk,S., Hunz,J., Iglebird,J., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levant,J., Lewis,L., Li,Z., Liu,J.,

COMMENT  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GICB  
 Center clone name: CH230-147P15  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 212164 bases at least Q40  
 Consensus quality: 213776 bases at least Q30  
 Consensus quality: 215303 bases at least Q20  
 Estimated insert size: 215883; sum-of-contigs estimation  
 Overall coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a working draft sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor;  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 217691: contig of 217691 bp in length.

FEATURES	Location/Qualifiers
source	1..217691
	/organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-147P15"
misc_feature	1..1410 /note="rgs-contig"
BASE COUNT	54935 a 50714 c 51140 g 59960 t 942 others
ORIGIN	

Query Match      100.0%    Score 18; DB 2; Length 217691;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

COMMENT

TITLE JOURNAL

Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 17, 2003 this sequence version replaced gi:28629983.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center : Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information -----  
 Center project name: L25016  
 Center clone name: 93\_G\_1

----- NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----

AC119882/c AC119882 Mus musculus clone RP24-33G1, WORKING DRAFT SEQUENCE, 10 unordered pieces.

ACCESSION AC119882.3

VERSION GT:30017888

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathini; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 244620)

AUTHORS Birren, B., Nusbaum, C. and Lander, E.

TITLE Mus musculus, clone RP24-33G1

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 244620)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barnet, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearle, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardna, S., Ginde, S., Gorz, S., Govertse, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Huine, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoocky, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, D., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihowa, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Polliara, V., Raymond, C., Rettar, R., Rieback, M., Riley, R., Rize, C., Rogov, P., Roman, J., Rosett, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trisilow, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

FEATURES

source 1..244620

organism="Mus musculus"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP24-33G1"  
 /clone\_lib="RPCI-24 Male Mouse BAC"

misc\_feature 1..2154

/note="assembly\_fragment  
 clone\_end=SP6  
 vector\_side:left"

REFERENCE 3 (bases 1 to 244620)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Bastien, V., Bloom, T., Anderson, S., Arachchi, H.M., Barna, N., Barna, N.,



Query Match									
Best Local Similarity		Score 17;		DB 2;		Length 47588;			
Matches 17;		Conservative 0;		Pred. No. 2.5e+0;		Mismatches 0;		Indels 0; Gaps 0;	
*	12784:	contig of 666 bp in length	*	41910	42009;	gap of 100 bp			
*	12884:	gap of 100 bp	*	42010	42696;	contig of 687 bp in length			
*	13554:	contig of 670 bp in length	*	*	42697	42796;	gap of 100 bp		
*	13655:	gap of 100 bp	*	*	42797	43463;	contig of 667 bp in length		
*	14420:	contig of 766 bp in length	*	*	43464	43563;	gap of 100 bp		
*	14521:	gap of 100 bp	*	*	43564	44224;	contig of 661 bp in length		
*	15213:	contig of 693 bp in length	*	*	44225	44324;	gap of 100 bp		
*	15214:	gap of 100 bp	*	*	44325	44985;	contig of 661 bp in length		
*	15314:	contig of 675 bp in length	*	*	44986	45085;	gap of 100 bp		
*	15988:	gap of 100 bp	*	*	45086	45862;	contig of 777 bp in length		
*	16088:	contig of 664 bp in length	*	*	45863	45962;	gap of 100 bp		
*	16752:	gap of 100 bp	*	*	45963	46732;	contig of 770 bp in length		
*	16753:	contig of 673 bp in length	*	*					
*	16785:	gap of 100 bp	*	*					
*	17626:	contig of 673 bp in length	*	*					
*	18299:	gap of 100 bp	*	*					
*	18399:	contig of 695 bp in length	*	*					
*	19094:	gap of 100 bp	*	*					
*	19194:	contig of 760 bp in length	*	*					
*	19953:	gap of 100 bp	*	*					
*	19954:	contig of 100 bp	*	*					
*	20054:	contig of 660 bp in length	*	*					
*	20714:	gap of 100 bp	*	*					
*	20814:	contig of 667 bp in length	*	*					
*	21481:	gap of 100 bp	*	*					
*	21581:	contig of 758 bp in length	*	*					
*	22339:	gap of 100 bp	*	*					
*	22439:	contig of 685 bp in length	*	*					
*	23124:	gap of 100 bp	*	*					
*	23224:	contig of 682 bp in length	*	*					
*	23906:	gap of 100 bp	*	*					
*	24006:	contig of 701 bp in length	*	*					
*	24707:	gap of 100 bp	*	*					
*	24807:	contig of 669 bp in length	*	*					
*	25476:	gap of 100 bp	*	*					
*	25576:	contig of 700 bp in length	*	*					
*	26276:	gap of 100 bp	*	*					
*	26376:	contig of 817 bp in length	*	*					
*	27193:	gap of 100 bp	*	*					
*	27293:	contig of 748 bp in length	*	*					
*	28041:	gap of 100 bp	*	*					
*	28891:	contig of 751 bp in length	*	*					
*	28992:	gap of 100 bp	*	*					
*	29649:	contig of 657 bp in length	*	*					
*	29749:	gap of 100 bp	*	*					
*	30525:	contig of 777 bp in length	*	*					
*	30526:	gap of 100 bp	*	*					
*	31393:	contig of 768 bp in length	*	*					
*	31394:	gap of 100 bp	*	*					
*	31494:	contig of 698 bp in length	*	*					
*	32291:	gap of 100 bp	*	*					
*	32192:	contig of 655 bp in length	*	*					
*	32292:	gap of 100 bp	*	*					
*	32947:	contig of 684 bp in length	*	*					
*	33047:	gap of 100 bp	*	*					
*	33731:	contig of 672 bp in length	*	*					
*	33831:	gap of 100 bp	*	*					
*	34503:	contig of 100 bp	*	*					
*	34603:	contig of 780 bp in length	*	*					
*	35383:	gap of 100 bp	*	*					
*	35483:	contig of 671 bp in length	*	*					
*	36155:	gap of 100 bp	*	*					
*	36255:	contig of 697 bp in length	*	*					
*	36952:	gap of 100 bp	*	*					
*	37051:	contig of 673 bp in length	*	*					
*	37725:	gap of 100 bp	*	*					
*	37825:	contig of 789 bp in length	*	*					
*	38495:	gap of 100 bp	*	*					
*	38595:	contig of 793 bp in length	*	*					
*	39389:	gap of 100 bp	*	*					
*	39489:	contig of 100 bp	*	*					
*	40277:	gap of 100 bp	*	*					
*	40378:	contig of 764 bp in length	*	*					
*	41141:	gap of 100 bp	*	*					
*	41242:	contig of 668 bp in length	*	*					

Query Match

Best Local Similarity 94.4%; Score 17;

Matches 17; Conservatve 0; Pred. No. 2.5e+0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAGGCAGGACAGA 17

Db 36306 GACAAGGCAGGACAGA 36322

Query Match

Best Local Similarity 100.0%; Score 100.0%; DB 2; Length 116608 bp; DNA linear

LOCUS AC121356 116608 bp; WORKING DRAFT SEQUENCE, 5 unordered pieces.

DEFINITION Mus musculus clone RP23-52D17, WORKING DRAFT SEQUENCE, 5 unordered pieces.

ACCESSION AC121356

VERSION AC121356\_3 GI:28975044

KEYWORDS HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Mus musculus (house mouse)

ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 116608)

AUTHORS Birren,B., Nusbaum,C., and Lander,E.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 116608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,E., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choppel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,W., Gage,D., Goto,S., Goyette,M., Graham,L., Galegan,J., Gardyna,S., Ginde,S., Gord,S., Goyle,M., Hulme,W., Iliev,I., Grand-Pierre,N., Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,T., LaRocke,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McErath,K., McErath,J., Meldrim,J., Menes,L., Mihova,T., Milenga,V., Murphy,T., Nayor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rose,C., Rogov,P., Roman,J., Rosetini,M., Roy,A., Santos,R., Schaefer,S., Schupack,R., Seaman,S., Severi,M., Spender-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wymond,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A., and Zody,M.

JOURNAL Direct Submission

RESEARCHER (17-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 116608)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,E., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cooke,P., Corum,B., Doolittle,K., Dorris,L., Erickson,J., Faro,S., Diaz,J.S., Dodge,S., Dooley,K., Horton,L., Iliev,I., Johnson,R., Jones,C., Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Perreira,P., Fitzgerald,W., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafes,N., Hafez,E., Haug,J., Hulme,W., Hall,J., Horton,L., Iliev,I., Hulme,W., Iliev,I., Jones,C., Jones,R.

Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., MacLean, C., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., McCarthy, M., McDonald, P., Major, J., Manning, J., Matthews, C., McConaughay, M., Meldrum, J., Menous, L., Mihova, T., Mileng, V., Naylor, J., Nguyen, C., Nicoll, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Rettig, R., Rogov, P., Ronan, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tasseff, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Vieil, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (16-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

**COMMENT** On Mar 16, 2003 this sequence version replaced gi|25901114.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

**Center:** Whitehead Institute/ MIT Center for Genome Research  
**Center code:** WIBR  
**Web site:** <http://www-seq.wi.mit.edu>  
**Contact:** [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

**Project Information**

**Center project name:** L23269  
**Center clone name:** 52-D-17

**Summary Statistics**

**Sequencing vector:** Plasmid; n/a; 100% of reads  
**Chemistry:** Dye-terminator Big Dye; 100% of reads  
**Assembly program:** Phrap; version 0.96031  
**Consensus quality:** 115769 bases at least Q30  
**Consensus quality:** 115895 bases at least Q30  
**Consensus quality:** 116130 bases at least Q20  
**Insert size:** 15000bp; agarose-fp  
**Insert size:** 116208; sum-of-contigs  
**Quality coverage:** 10.2 in Q20 bases; agarose-fp  
**Quality coverage:** 13.1 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 23731: contig of 23731 bp in length  
 \* 23832: 23831: gap of 100 bp  
 \* 241019: 241019: contig of 17168 bp in length  
 \* 41120: 41119: gap of 100 bp  
 \* 60788: 60788: contig of 19669 bp in length  
 \* 60888: 60888: gap of 100 bp  
 \* 60889: 97152: contig of 36264 bp in length  
 \* 97152: 97253: gap of 100 bp  
 \* 97253: 116608: contig of 19356 bp in length.

**FEATURES**

**Source**

**misc\_feature**

1. .116608  
 /note="assembly\_fragment"  
 /clone\_end\_SP6  
 /vector\_Side:left"  
 23832. 41019  
 /note="assembly\_fragment"  
 /db\_xref="taxon:10090"  
 /clone\_id="RP23-52D17"

**misc\_feature**

41120. 60788  
 /note="assembly\_fragment"  
 /note="assembly\_fragment"  
 60889. 97152  
 /note="assembly\_fragment"

Query Match 94.4%; Score 17; DB 2; Length 116608;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACAGGGCAGGACAGA 17  
 Db 36482 GACAAGGGGAGGAGA 36498

**RESULT** 10  
**ACCESSION** AC144886/C  
**LOCUS** AC144886\_1 GI:31044305  
**VERSION** AC144886\_1 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
**KEYWORDS** Bos taurus (cow)  
**ORGANISM** Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Bivalvia; Mammalia; Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovine; Bos.  
 1 (bases 1 to 151613)  
**REFERENCE**  
**AUTHORS** Antonellis, A., Ayelle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Granite, S., Gruen, X., Chu, G., Coleman, H., Colemen, H., Engle, J., Hansen, J., Ho, S.-L., Hu, P., Gupta, S., Haghghi, P., Han, J., Karlin, E., Kwon, P., Lee-Lin, S., Q., Hurle, B., Idol, J.R., Maduro, Q.L., Maruyama, F.H., Masilio, C., Legspis, R., Maduro, Q.L., McDowell, J., Paganigan, C., Pearson, R., Portnoy, M.E., Mastri, B., Mastri, B., McDowell, J., Paganigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Reddix-Pugue, N., Schueler, R., Shah, R., Sison, C., Stantripp, S., Thomas, J.W., Thomas, P.J., Tsipouri, V., Vogt, J.L., Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.

**COMMENT** Unpublished  
 NISC Comparative Sequencing Initiative  
 Direct Submission  
 Submitted (23-MAY-2003) NIH Intramural Sequencing Center, 8717 Grovermont Circle, Gaithersburg, MD 20877, USA  
**COMMENT** ----- Genome Center  
**Center** NIH Intramural Sequencing Center  
**Center code** NISC  
**Web site** <http://www.nisc.nih.gov>  
**Contact** nisc.zoo@nihgr.nih.gov  
 ----- Project Information  
**Center project name** djc  
**Center clone name** 147291  
 ----- Summary Statistics  
**Sequencing vector:** Plasmid; n/a; 100% of reads  
**Chemistry:** Dye-terminator Big Dye; 100% of reads  
**Assembly program:** Phrap; version 0.990319  
**Consensus quality:** 14535 bases at least Q40  
**Consensus quality:** 146470 bases at least Q30  
**Consensus quality:** 147291 bases at least Q20  
**Insert size:** 120000; agarose-EP  
 ----- Quality coverage: 11.18x in Q20 bases; agarose-fp  
 ----- Quality coverage: 8.93x in Q20 bases; sum-of-contigs  
 ----- Quality coverage: 11.18x in Q20 bases; agarose-fp  
 ----- Quality coverage: 8.93x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence

* as soon as it is available and the accession number will be preserved.	1	2844: contig of 2844 bp in length
*	2845	2944: gap of unknown length
*	2945	5124: contig of 2180 bp in length
*	5124	5224: gap of unknown length
*	5225	9700: contig of 4536 bp in length
*	9761	9860: gap of unknown length
*	9861	13071: contig of 3211 bp in length
*	13072	13171: gap of unknown length
*	13172	17792: contig of 4621 bp in length
*	17793	17894: gap of unknown length
*	17893	24116: contig of 6224 bp in length
*	24217	24218: gap of unknown length
*	30530	30529: contig of 6313 bp in length
*	30630	30639: gap of unknown length
*	42067	42067: contig of 11438 bp in length
*	42068	42167: gap of unknown length
*	42168	51575: contig of 9408 bp in length
*	51576	51675: gap of unknown length
*	51676	62323: contig of 10654 bp in length
*	62330	62429: gap of unknown length
*	62430	72849: contig of 10420 bp in length
*	72850	72943: gap of unknown length
*	72950	86333: contig of 13414 bp in length
*	86364	86463: gap of unknown length
*	86464	99995: contig of 13532 bp in length
*	99995	100095: gap of unknown length
*	100096	122705: contig of 22610 bp in length
*	122706	122806: gap of unknown length
*	122806	151613: contig of 28808 bp in length

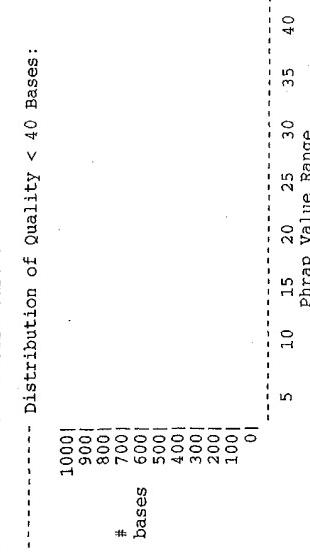
Location	Qualifiers	Source	SE COUNT
1..151613			
	/organism="Bos taurus"		
	/mol_type="genomic DNA"		
	/db_xref=taxon:9013		
	/clone="RP42-563J22"		
	/clone_id="RP42"		
1..2844			
	/note="assembly_fragment"		
2945..5124			
	/note="assembly_fragment"		
5225..9760			
	/note="assembly_fragment"		
9861..13071			
	/note="assembly_fragment"		
13172..17792			
	/note="assembly_fragment"		
17893..24116			
	/note="assembly_fragment"		
24217..30529			
	/note="assembly_fragment"		
30630..42067			
	/note="assembly_fragment"		
42168..551575			
	/note="assembly_fragment"		
51676..62329			
	/note="assembly_fragment"		
62430..72849			
	/note="assembly_fragment"		
72950..86363			
	/note="assembly_fragment"		
86464..99995			
	/note="assembly_fragment"		
	clone_end:SP6		
	vector_side:right,"		
100096..122705			
	/note="assembly_fragment"		
122806..151613			
	/note="assembly_fragment"		
	clone_end:T7		
	vector_side:right,"		
31102 a	447433 c 42876 o 31470 t 1402 others		

ORIGIN	Query	Best I	
	Matche		
QY			
Db			
RESULT	1		
AC027761			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORD			
SOURCE			
ORGANI			
REFERENCE			
AUTHOR			
TITLE			
JOURNAL			
REFERENCE			
AUTHOR			
TITLE			
JOURNAL			
REFERENCE			
AUTHOR			
TITLE			
JOURNAL			
REFERENCE			
AUTHOR			
TITLE			
JOURNAL			
REFERENCE			
AUTHOR			
TITLE			
JOURNAL			
COMMENT			

**ANNOTATION OF FEATURES:**  
 STSs are identified using ePCR (Genome Res. 7:541-550).  
 STSs are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse Sequences.  
 Genes and gene sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3399-3402) similarity ( $\text{expect} < 1e-34$ ) to EST  
 and cDNA sequences in Unigene. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintain sequence continuity  
 across the splice junctions. Sequences that are not identical  
 matches are annotated as similar.

**SEQUENCING READ COVERAGE:** Attempts are made to complete double  
 stranded sequence for all regions. All sequence is completed to a  
 standard of coverage with a minimum of 3 reads with no ambiguities  
 if the sequence coverage for a region does not meet this standard,  
 it is indicated in the annotation as Low Coverage. Low coverage  
 linkages are verified by PCR product size verification or  
 verification of forward and reverse reads from clones which span  
 the low coverage area.

**QUALITY OF INDIVIDUAL BASES:** This sequence meets stringent quality  
 standards - estimated average error rate is less than 1 per 10,000  
 bases using the Consed quality parameters. Regions that do not  
 meet this requirement are annotated as Low Quality.



```

Query Match      94.4%; Score 17; DB 10; Length 163475;
Best Local Similarity 100.0%; Fred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0;
> 2 ACAGGGCCGACGAC 18

```

20

**RESULT 12**  
**AC145188/c** AC145188 173362 bp DNA linear HTG 12-JUN-2003  
**DEFINITION** Bos taurus clone RP42-483Q023, WORKING DRAFT SEQUENCE, 16 unordered pieces.  
**ORGANISM** Mammalia; Bovidae; Bovinae; Bos.  
**ACCESSION** AC145188.1 GI:31621331  
**VERSION** AC145188.1 to 173362  
**KEYWORD** HTG; HTGS\_PHASEI; HTGS\_DRAFT.  
**SOURCE** Bos taurus (cow)  
**BITS** Bos tauri  
**PIECES** Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.  
**REFERENCE** Antonisse,A., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K., Chu,G., Coleman,B., Coley,J., Engle,J., Granite,S., Guan,X., Gupta,J., Hachighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Leric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Marquilles,E.H., Mastellino,C., Maskeri,B., McDowell,J., Pagurigan,C., Pearson,R., Portnoy,M.E., Prasad,A., Readiix-Dugue,N., Schandier,K., Schueler,M.G., Shah,K., Sison,C., Stantrapop,S., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,J.W., Young,L., and Green,E.D.  
**TITLE** NTSC Comparative Sequencing Initiative  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 173362)  
**AUTHORS** Green,F.D.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (12-JUN-2003) NIH Intramural Sequencing Center, 8717  
**COMMENT** Groenvont Circle, Gaithersburg, MD 20877, USA  
**COMMENT** Center: NIH Intramural Sequencing Center  
**COMMENT** Center code: NISC  
**COMMENT** Web site: <http://www.niss.nih.gov>  
**COMMENT** Contact: nisc.zoonhangr.nih.gov  
**COMMENT** Project Information  
**Center** project name: djd  
**Center** clone name: 483Q023  
**COMMENT** Summary Statistics  
**Sequencing vector:** plasmid; **n/a;** 100% of reads  
**Chemistry:** Dy-terminator Big Dye; 100% of reads  
**Assembly program:** Phrap; version 0.990319  
**Consensus Quality:** 168024 bases at least Q40  
**Consensus Quality:** 169046 bases at least Q30  
**Consensus Quality:** 168811 bases at least Q20  
**Insert size:** 158000; agarose-**fp**  
**Insert size:** 171862; sum-of-contigs  
**Quality coverage:** 11.06x in Q20 bases; agarose-**fp**  
**Quality coverage:** 10.17x in Q20 bases; sum-of-contigs  
**NOTE:** This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
**1** 2487: contig of 2487 bp in length  
**2** 2587: gap of unknown length  
**3** 4812: contig of 2225 bp in length  
**4** 4913: gap of unknown length  
**5** 9311: contig of 4399 bp in length  
**6** 9411: gap of unknown length  
**7** 11916: contig of 2505 bp in length  
**8** 12016: gap of unknown length  
**9** 11917: 14880; contig of 2864 bp in length  
**10** 12117: 14981; gap of unknown length  
**11** 12193: 33233 a 52193 c 51596 g 34837 t 1513 others  
**BASE COUNT** 33233 a 52193 c 51596 g 34837 t 1513 others  
**ORIGIN** Query Match 94.4%; Score 17; DB 2; Length 173362;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0  
**Db** 35764 GACAGGGCAGGACAGA 17  
**Db** 35764 GACAGGGCAGGACAGA 35748.

RESULT 13  
AC138402/C AC138402 linear 196351 bp DNA 07-MAR-2003 HTG 07-MAR-2003  
LOCUS Mus musculus chromosome 14 clone RP24-263011 map 14, WORKING DRAFT  
DEFINITION SEQUENCE, 3 unordered pieces.  
ACCESSION AC138402  
VERSION AC138402.3 GI:28876093  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
ORGANISM Mus musculus (house mouse)  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 196351)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Unpublished  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 196351)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,  
Camarata,J., Chang,J., Chazarian,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Haftz,N.,  
Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Macdonald,P., Major,J.,  
Matthews,C., McCarthy,M., Medrano,J., Meneus,L., Minova,T.,  
Miengwa,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunkhahang,P., Pierre,N., Raymond,C., Reita,R.,  
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
stojanovic,N., Talama,J., Tasfaye,S., Theodore,J., Topham,K.,  
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,J., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission (29-DEC-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 196351)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Garunya,S.,  
Graham,L., Grand-Pierre,N., Haftz,N., Haggopian,D., Hagos,B.,  
Hall,J., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Neldrake,J., Meneus,L., Minova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,J., Peterson,K., Phunkhahang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Reita,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seeman,S., Severy,P.,  
Spencer,B., Stange-Thomann,N., Stolancic,N., Stubbs,M.,  
Talama,J., Tasfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Submitted (07-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Mar 7, 2003 this sequence version replaced gi:28461005.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

RESULT 14  
AC079082 AC079082 linear 205621 bp DNA 15-NOV-2002  
LOCUS Mus musculus chromosome 10 clone rp23-161b1 strain C57BL/6J,  
DEFINITION complete sequence.  
ACCESSION AC079082  
VERSION AC079082  
KEYWORDS HTG,  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE Ying,F. and Roe,B.A.  
AUTHORS  
TITLE

Center: Whitehead Institute, MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information

REFERENCE	Unpublished 2 (bases 1 to 205621)	JOURNAL	
AUTHORS	Ying F. and Roe, B.A.	AUTHORS	
TITLE	Direct Submission	TITLE	
JOURNAL	Submitted (18-AUG-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	JOURNAL	
REFERENCE	3 (bases 1 to 205621)	REFERENCE	
AUTHORS	Ying,F. and Roe,B.A.	AUTHORS	
TITLE	Direct Submission	TITLE	
JOURNAL	Submitted (04-JAY-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	JOURNAL	
REFERENCE	4 (bases 1 to 205621)	REFERENCE	
AUTHORS	Ying,F. and Roe,B.A.	AUTHORS	
TITLE	Direct Submission	TITLE	
JOURNAL	Submitted (07-MAY-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	JOURNAL	
REFERENCE	5 (bases 1 to 205621)	REFERENCE	
AUTHORS	Ying,F. and Roe,B.A.	AUTHORS	
TITLE	Direct Submission	TITLE	
JOURNAL	Submitted (15-NOV-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	JOURNAL	
COMMENT	On Nov 15, 2002 this sequence version replaced gi:20422347.	COMMENT	
FEATURES	Location/Qualifiers	FEATURES	
SOURCE	1..205621	SOURCE	
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/strain="C57BL/6J"		
	/db_xref="TAXON:10990"		
	/chromosome="10"		
	/clone_id="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"		
BASE COUNT	52129 a 53223 c 50420 g 49849 t	BASE COUNT	
ORIGIN		ORIGIN	
Query Match	94.4%; Score 17; DB 10; Length 205621;	Query Match	
Best Local Similarity	100.0%; Pred. No. 2.e+02;	Best Local Similarity	
Matches	17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches	
QY	2 ACAGGGCAGGACAGAG 18	QY	
Db	16542 ACAGGGCAGGACAGAG 16558	Db	
RESULT	15	RESULT	
AC132033	AC132033	AC132033	
LOCUS	217181 bp DNA linear HTG 19-NOV-2002	LOCUS	
DEFINITION	Rattus norvegicus clone CH230-127G19, WORKING DRAFT SEQUENCE, 3	DEFINITION	
ACCESSION	AC132033	ACCESSION	
VERSION	AC132033.4	VERSION	
KEYWORDS	GI:25091414	KEYWORDS	
SOURCE	HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	SOURCE	
ORGANISM	Rattus norvegicus	ORGANISM	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae;	
REFERENCE	1 (bases 1 to 217181)	REFERENCE	
AUTHORS	Muzny,D.Marie., Allen,H., Alsbrooks,S., Amin,A., Angulano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barrer,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blith,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,R., Cavazos,I., Ceasar,H., Center,A.,	AUTHORS	

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyne,M., Cree,A., D'Souza,B., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,J., Denzon,S., Derano,C., Ding,Y., Dinh,H., Drury,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,T., Eaves,K., Egan,A., Escott,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabis,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Gear,K., Gill,R., Grady,M., Guerra,W., Guyvara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hernandez,J., Harvey,K., Havlik,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodges,A., Hogue,M., Hollins,B., Howell,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolive,A., Karpfay,S., Kelly,S., Kelly,J., King,L., Kovar,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., Landon,P., Longacre,S., Lopez,J., Loveland,R., Loulseged,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindarine,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,B., Mawhinney,S., McLeod,M.P., McNeil,T.Z., Meinen,E., Milesavlevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasan,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nzakieleh,O., Okwounou,G., Olarumpungason,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,P., Poindexter,A., Popovic,D., Primus,E., Prill,L., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R.Q., Ruiz,S.J.J., Sanders,N., Saveri,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvarcbeyn,A., Sisson,I., Sitter,C.D., Smajis,D., Sned,A., Sodergren,E., Song,X.-Z., Sorel,R., Sosa,J., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wiedey,R., Woodin,R., Worley,K., Wright,D., Wright,R., Wu,J., Yakuub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Smith,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRX  
 Center clone name: CH230-127G19  
 ----- Summary Statistics  
 Assembly program: PIRAP; version 0.990329  
 Consensus quality: 21280 bases at least Q40  
 Consensus quality: 21361 bases at least Q30  
 Consensus quality: 21427 bases at least Q20  
 Estimated insert size: 21841; sum-of-contigs estimation  
 Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

---

- \* NOTE: Estimated insert size may differ from sequence length.  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).
- \* NOTE: This is a working draft sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
- \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
- \* 211237: contig of 211237 bp in length
- \* 211238 111337: gap of unknown length
- \* 211338 211612: contig of 3275 bp in length
- \* 214613 214712: gap of unknown length
- \* 214713 217181: Contig of 2469 bp in length.

FEATURES

Source	1..217181	Location/Qualifiers
		/organism="Rattus norvegicus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10116"
		/clone="CH230-127G19"

BASE COUNT

ORIGIN	60118 a	46007 C	47603 g	60919 t	2534 others
--------	---------	---------	---------	---------	-------------

Query Match 94.4%; Score 17; DB 2; Length 217181;  
 Best Local Similarity 100.0%; Fred. No. 2.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	ACAGGGCAGGACAGAG 18
Db	12734	ACAGGGCAGGACAGAG 12750

Search completed: August 27, 2003, 18:41:51  
 Job time : 542.538 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:42:00 ; Search time 173.269 Seconds  
(without alignments)  
450.193 Million cell updates/sec

Title: US-10-085-612-2

Perfect score: 34  
Sequence: 1 cgtattttgtactggctgaggctgcagcccg 34

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Maximum Match 100%  
Listing first 45 summaries

Database :

```
1: /cgn2_6/ptodata/2/pdbna/US07_PUBCOMB.seq;*
2: /cgn2_6/ptodata/2/pdbna/US06_NEW_PUB.seq;*
3: /cgn2_6/ptodata/2/pdbna/US05_PUBCOMB.seq;*
4: /cgn2_6/ptodata/2/pdbna/US05_NEW_PUB.seq;*
5: /cgn2_6/ptodata/2/pdbna/US04_PUB.seq;*
6: /cgn2_6/ptodata/2/pdbna/PCUTS_PUBCOMB.seq;*
7: /cgn2_6/ptodata/2/pdbna/US08_NEW_PUB.seq;*
8: /cgn2_6/ptodata/2/pdbna/US08_PUBCOMB.seq;*
9: /cgn2_6/ptodata/2/pdbna/US09_PUBCOMB.seq;*
10: /cgn2_6/ptodata/2/pdbna/US09B_PUBCOMB.seq;*
11: /cgn2_6/ptodata/2/pdbna/US09C_PUBCOMB.seq;*
12: /cgn2_6/ptodata/2/pdbna/US09_NEW_PUB.seq;*
13: /cgn2_6/ptodata/2/pdbna/US10A_PUBCOMB.seq;*
14: /cgn2_6/ptodata/2/pdbna/US10B_PUBCOMB.seq;*
15: /cgn2_6/ptodata/2/pdbna/US10C_PUBCOMB.seq;*
16: /cgn2_6/ptodata/2/pdbna/US60_NEW_PUB.seq;*
17: /cgn2_6/ptodata/2/pdbna/US60_PUBCOMB.seq;*
```

Published Applications NA:  
1: /cgn2\_6/ptodata/2/pdbna/US07\_PUBCOMB.seq;\*

2: /cgn2\_6/ptodata/2/pdbna/US06\_NEW\_PUB.seq;\*

3: /cgn2\_6/ptodata/2/pdbna/US05\_PUBCOMB.seq;\*

4: /cgn2\_6/ptodata/2/pdbna/US05\_NEW\_PUB.seq;\*

5: /cgn2\_6/ptodata/2/pdbna/US04\_PUB.seq;\*

6: /cgn2\_6/ptodata/2/pdbna/PCUTS\_PUBCOMB.seq;\*

7: /cgn2\_6/ptodata/2/pdbna/US08\_NEW\_PUB.seq;\*

8: /cgn2\_6/ptodata/2/pdbna/US08\_PUBCOMB.seq;\*

9: /cgn2\_6/ptodata/2/pdbna/US09\_PUBCOMB.seq;\*

10: /cgn2\_6/ptodata/2/pdbna/US09B\_PUBCOMB.seq;\*

11: /cgn2\_6/ptodata/2/pdbna/US09C\_PUBCOMB.seq;\*

12: /cgn2\_6/ptodata/2/pdbna/US09\_NEW\_PUB.seq;\*

13: /cgn2\_6/ptodata/2/pdbna/US10A\_PUBCOMB.seq;\*

14: /cgn2\_6/ptodata/2/pdbna/US10B\_PUBCOMB.seq;\*

15: /cgn2\_6/ptodata/2/pdbna/US10C\_PUBCOMB.seq;\*

16: /cgn2\_6/ptodata/2/pdbna/US60\_NEW\_PUB.seq;\*

17: /cgn2\_6/ptodata/2/pdbna/US60\_PUBCOMB.seq;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Description	Length	DB ID
1	34	100.0	34	14	US-10-085-612-2
2	33	97.1	34	14	US-10-085-612-6
3	33	97.1	1254	14	US-10-085-612-4
4	22.8	67.1	1012	10	US-09-957-997-4
5	22.8	67.1	1345	14	US-10-146-515-3
6	22.8	67.1	1345	14	US-10-083-612-3
7	22.8	67.1	1186	10	US-09-955-997-1
8	22.8	67.1	13035	12	US-10-121-960C-14
9	22.8	67.1	1585	12	US-10-121-960C-17
c 10	21.6	63.5	481	13	US-10-027-632-25780
c 11	21.4	62.9	599	13	US-10-027-632-25947
c 12	20.6	60.6	30	14	US-10-215-132-41
c 13	20.2	59.4	470	11	US-09-918-955-2594
c 14	20.2	59.4	672	13	US-10-027-632-233693
c 15	20.2	59.4	994	11	US-09-934-455-189
c 16	58.8	62.7	9	us-09-770-149-726	

Sequence 322, App  
Sequence 11, App  
Sequence 204755, App  
Sequence 3264, App  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 1908, App  
Sequence 3727, App  
Sequence 278, App  
Sequence 255972, App  
Sequence 158, App  
Sequence 255971, App  
Sequence 25597, App  
Sequence 334, App  
Sequence 2824, App  
Sequence 3651, App  
Sequence 29798, App  
Sequence 8079, App  
Sequence 7049, App  
Sequence 50, Appli  
Sequence 570, Appli  
Sequence 571, App  
Sequence 593-501-570  
Sequence 513, App  
Sequence 29785-300-513  
Sequence 26, Appli  
Sequence 3, Appli  
Sequence 486, App  
Sequence 5486, App  
Sequence 3210, App  
Sequence 4985, App  
Sequence 322496, App

## ALIGNMENTS

RESULT 1  
US-10-085-612-2  
; Sequence 2, Application US/10085612  
; GENERAL INFORMATION  
; PUBLICATION NO. US2010030096251A1  
; APPLICANT: Guida, Marco  
; APPLICANT: Hall, Jeff  
; APPLICANT: Petros, William  
; APPLICANT: Vredenburg, James  
; APPLICANT: Colvin, Oliver  
; APPLICANT: Marks, Jeffrey  
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceutical Compounds Therefor  
; FILE REFERENCE: 439-5-CI  
; CURRENT FILING DATE: 2002-02-26  
; PRIORITY APPLICATION NUMBER: 09/144,367  
; PRIORITY FILING DATE: 1998-08-31  
; PRIORITY APPLICATION NUMBER: 60/271,630  
; PRIORITY FILING DATE: 2001-02-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
; LENGTH: 34  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-085-612-2  
Query Match 100.0% ; Score 34; DB 14; Length 34;  
Best Local Similarity 100.0%; Mismatches 0; Indels 0; Gaps 0;  
Matches 34; Conservative 0; Organism US-10-085-612-2  
Qy 1 CGATCTCTGTACTCGCTGAGTGCACCCCG 34  
Db 1 CGATCTCTGTACTCGCTGAGTGCACCCCG 34

Sequence 6, Application US/10085612  
 Publication No. US2003009625A1  
 GENERAL INFORMATION:  
 APPLICANT: Guida, Marco  
 APPLICANT: Hall, Jeff  
 APPLICANT: Petros, William  
 APPLICANT: Vredenburgh, James  
 APPLICANT: Colvin, Oliver  
 APPLICANT: Marks, Jeffrey  
 TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals  
 TITLE OF INVENTION: Compositions Therefor  
 FILE REFERENCE: 4389-5-CL  
 CURRENT APPLICATION NUMBER: US10/085,612  
 CURRENT FILING DATE: 2002-02-26  
 PRIOR APPLICATION NUMBER: 09/144,367  
 PRIOR FILING DATE: 1998-08-31  
 PRIOR APPLICATION NUMBER: 60/271,630  
 PRIOR FILING DATE: 2001-02-26  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO: 6  
 LENGTH: 34  
 TYPE: DNA  
 ORGANISM: Homo sapiens

US-10-085-612-6

Query Match 97.1%; Score 33; DB 14; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 0.00034; Indels 0; Gaps 0;  
 Matches 33; Conservative 0; Mismatches 0; Location: (0) ... (0)

QY 1 CGATCTTGCTACTGGTGAAGCTGAGCTGACCCCC 33  
 Db 1 CGATCTTGCTACTGGTGAAGCTGAGCTGACCCCC 33

RESULT 3  
 US-10-085-612-4

Sequence 4, Application US/10085612  
 Publication No. US2003009625A1  
 GENERAL INFORMATION:  
 APPLICANT: Guida, Marco  
 APPLICANT: Hall, Jeff  
 APPLICANT: Petros, William  
 APPLICANT: Vredenburgh, James  
 APPLICANT: Colvin, Oliver  
 APPLICANT: Marks, Jeffrey  
 TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals  
 TITLE OF INVENTION: Compositions Therefor  
 FILE REFERENCE: 4389-5-CL  
 CURRENT APPLICATION NUMBER: US10/085,612  
 CURRENT FILING DATE: 2002-02-26  
 PRIOR APPLICATION NUMBER: 09/144,367  
 PRIOR FILING DATE: 1998-08-31  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO: 4  
 LENGTH: 1254  
 TYPE: DNA  
 ORGANISM: Homo sapiens

US-10-085-612-4

Query Match 97.1%; Score 33; DB 14; Length 1254;  
 Best Local Similarity 100.0%; Pred. No. 0.00049; Indels 0; Gaps 0;  
 Matches 33; Conservative 0; Mismatches 0; Location: (0) ... (0)

QY 1 CGATCTTGCTACTGGTGAAGCTGACCCCC 33  
 Db 1 CGATCTTGCTACTGGTGAAGCTGACCCCC 1036

RESULT 4

APPLICANT: Vredenburgh, James  
; APPLICANT: Colvin, Oliver  
; APPLICANT: Marks, Jeffrey  
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals  
; TITLE OF INVENTION: Compositions Therefor  
; FILE REFERENCE: 4389-5-C1  
; CURRENT APPLICATION NUMBER: US/10/085,612  
; CURRENT FILING DATE: 2002-02-16  
; PRIOR APPLICATION NUMBER: 09/144,367  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/271,630  
; PRIOR FILING DATE: 2001-02-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO: 3  
; LENGTH: 1345  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-085-612-3

Query Match 67.1%; Score 22.8; DB 14; Length 1345;  
Best Local Similarity 92.3%; Pred. No. 7.2; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 2; Software: PatentIn Ver. 3.0  
Db 1035 TGCTACTGGCTGAGGTGGCAGCCCCG 34

---

RESULT 7  
US-09-957-997-1  
Sequence 1, Application US/09957997  
; Patent No. US2002010915A1  
; GENERAL INFORMATION:  
; APPLICANT: Berklund, Anders  
; APPLICANT: Bertilsson, Gran  
; APPLICANT: Blomquist, Patrik  
; FILE OF INVENTION: PROMOTER SEQUENCES  
; FILE REFERENCE: 13425-046001  
; CURRENT APPLICATION NUMBER: US/09/957,997  
; CURRENT FILING DATE: 2001-09-21  
; EARLIER APPLICATION NUMBER: SE 0003393-6  
; EARLIER FILING DATE: 2000-09-22  
; EARLIER APPLICATION NUMBER: 60/238,895  
; EARLIER FILING DATE: 2000-10-10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 1  
; LENGTH: 11186  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-957-997-1

Query Match 67.1%; Score 22.8; DB 10; Length 11186;  
Best Local Similarity 92.3%; Pred. No. 8.9; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 2; Software: PatentIn Ver. 3.0  
Db 11064 TGCTACTGGCTGAGGTGGCAGCCCCG 11089

---

RESULT 8  
US-10-121-960C-14  
Sequence 14, Application US/10121960C  
; Publication No. US20030145341A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHANG, Weisheng  
; APPLICANT: CONFAG, Pamela  
; APPLICANT: PURCHIO, Anthony  
; APPLICANT: HASHIMA, Sandy  
; APPLICANT: NAVOTKA, Kevin  
; APPLICANT: MA, Shirley  
; APPLICANT: HOSHINO, Keiji  
; APPLICANT: NAWATKA, Keviin  
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF MOUSE AND HUMAN  
; TITLE OF INVENTION: CYTOCHROME EXPRESSION  
; FILE REFERENCE: 9400-0014 / PXE-014 US  
; CURRENT APPLICATION NUMBER: US/10/121,960C  
; CURRENT FILING DATE: 2002-04-11  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 17  
; LENGTH: 15185  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: CYP3A4-luc transgene  
US-10-121-960C-17

Query Match 67.1%; Score 22.8; DB 12; Length 15185;  
Best Local Similarity 92.3%; Pred. No. 9.2; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 2; Software: PatentIn Ver. 3.0  
Db 12860 TGCTACTGGCTGAGTCAGCCCTG 12885

---

RESULT 10  
US-10-027-632-72580  
Sequence 7, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827-129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIORITY APPLICATION NUMBER: US 60/218,006  
; PRIORITY FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; GENERAL INFORMATION:  
; APPLICANT: Eric B. Kniec  
; TITLE OF INVENTION: Composition and methods for the prevention and treatment of Huntington's disease  
; FILE REFERENCE: Napo-10  
; CURRENT APPLICATION NUMBER: US/10/215,432  
; CURRENT FILING DATE: 2002-11-19  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 41  
; LENGTH: 30  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Converted HD sequence  
US-10-027-632-72580

Query Match 62.9%; Score 21.4; DB 13; Length 599;  
Best Local Similarity 80.6%; Pred. No. 25;  
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 ATCTTGCTACTGGCGAGCC 33  
Db 48 ATCTTACACCTGAGTCAGCACC 18

RESULT 14  
US-10-027-632-285947/C  
; Sequence 28594, Application US/10/027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827\_129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 285947  
; LENGTH: 599  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-285947

Query Match 62.9%; Score 21.4; DB 13; Length 599;  
Best Local Similarity 80.6%; Pred. No. 25;  
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 ATCTTGCTACTGGCGAGCC 33  
Db 48 ATCTTACACCTGAGTCAGCACC 18

RESULT 15  
US-09-918-995-24594  
; Sequence 24594, Application US/09/918995  
; GENERAL INFORMATION:  
; APPLICANT: Hisq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO: 24594  
; LENGTH: 470  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(470)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-24594

Query Match 59.4%; Score 20.2; DB 11; Length 470;  
Best Local Similarity 75.8%; Pred. No. 74;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 CGATTTGTGATGGCACTGGAGGCC 33  
Db 126 CGSTCCATATGCTGGCTGCATCTCACACC 158

RESULT 16  
US-10-027-632-253693  
; Sequence 253693, Application US/10/027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827\_129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 285947  
; LENGTH: 599  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-253693

Query Match 62.9%; Score 21.4; DB 13; Length 599;  
Best Local Similarity 80.6%; Pred. No. 25;  
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 ATCTTGCTACTGGCGAGCC 33  
Db 48 ATCTTACACCTGAGTCAGCACC 18

RESULT 17  
US-10-027-632-253693  
; Sequence 253693, Application US/10/027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827\_129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 253693  
; LENGTH: 672  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-027-632-253693

Query Match 59.4%; Score 20.2; DB 13; Length 672;  
Best Local Similarity 75.8%; Pred. No. 77; Mismatches 0; Indels 0; Gaps 0;

---

QY 2 GATTCTTGACTGCTGGAGCTGGCG 34  
Db 548 GAGCCUGGTCCTCGGTGCAATGAGCCAG 580

RESULT 15  
US-09-934-455-489  
; Sequence 489, Application US/09934455  
; Publication No. US20030121070A1  
; GENERAL INFORMATION:  
; APPLICANT: Adam, Luc  
; APPLICANT: Cirelman, Robert  
; APPLICANT: Dubell, Arnold  
; APPLICANT: Beard, Jacqueline  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Keddie, James  
; APPLICANT: Pilgrim, Marsha  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Reuber, Lynne  
; APPLICANT: Riechmann, Jose Luis  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Pineda, Omaira  
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV  
; FILE REFERENCE: MBI-0025  
; CURRENT APPLICATION NUMBER: US/09/934,455  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/227439  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: MBI-0022  
; PRIOR FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: MBI-0023  
; PRIOR FILING DATE: 2001-04-17  
; NUMBER OF SEQ ID NOS: 516  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 489  
; LENGTH: 994  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(711)  
; OTHER INFORMATION: G2547

Query Match 59.4%; Score 20.2; DB 11; Length 994;  
Best Local Similarity 88.0%; Pred. No. 80; Mismatches 3; Indels 0; Gaps 0;

US-09-934-455-489